

CALL NO.

CA1

ML314

-83D25

GOVT

Employment and
Immigration Canada

Emploi et
Immigration Canada

Government
Publications



3 1761 11767832 6

The Canadian Occupational Projection System

DETERMINING THE OPTIMAL
LEVEL OF INDUSTRY
DISAGGREGATION -
PROGRESS REPORT

AUTHOR: NORM LECKIE

STRATEGIC POLICY & PLANNING
LABOUR MARKET OUTLOOK AND
STRUCTURAL ANALYSIS DIRECTORATE

Canada



Employment and
Immigration Canada

Emploi et
Immigration Canada

Government
Publications

CAI
MI 314
83D25

The Canadian Occupational Projection System

DETERMINING THE OPTIMAL
LEVEL OF INDUSTRY
DISAGGREGATION -
PROGRESS REPORT

AUTHOR: NORM LECKIE

STRATEGIC POLICY & PLANNING
LABOUR MARKET OUTLOOK AND
STRUCTURAL ANALYSIS DIRECTORATE



CAI
MI314
-83D25

(LX)

THE CANADIAN OCCUPATIONAL
PROJECTION SYSTEM
COPS


DETERMINING THE OPTIMAL
LEVEL OF INDUSTRY
DISAGGREGATION -
PROGRESS REPORT

AUTHOR: NORM LECKIE

STRATEGIC POLICY & PLANNING
LABOUR MARKET OUTLOOK AND
STRUCTURAL ANALYSIS DIRECTORATE

NOVEMBER 1983

WH-3-408B



Digitized by the Internet Archive
in 2023 with funding from
University of Toronto



<https://archive.org/details/31761117678326>

Using Cluster Analysis to Determine the
Optimal Level of Industry Disaggregation:
Progress Report

Introduction

In this paper, we present the results of an attempt to arrive at a workable set of industries in an empirically sound manner.¹ Till now, COPS has made use of the 69 industries used by its predecessor, COFOR. Though such a breakdown was undoubtedly chosen for logical reasons, those reasons were never put down on paper. In this study we strive to do just that. Since occupations are our primary concern in COPS, our goal here is to achieve an optimal disaggregation of industries based on an empirical comparison of their respective occupational distributions.

In the first section, we outline the steps involved in preparing the original Census file in a form suitable for empirical analysis. In the next part is introduced a short exposé of the technique we will be using -- cluster analysis -- and its advantages. In section III, some methodological and statistical reservations are introduced. In the following section, the results of applying cluster analysis to the occupational coefficients of the Census industries are presented and discussed. There is a brief conclusion at the end of this report.

¹ In essence, this is a follow-up to the initial proposal, "The Level of Industry Disaggregation in COPS: A Discussion Paper", March 1983, presented at the April '83 Technical Conference.

I. Preparing the Data for Empirical Analysis

On our original combined Census tape, there are records only for each detailed occupation that exists in 1981 within each 3-digit SIC industry (for those occupations not required by the industry there is no record). On those records is contained the 1971 and 1981 employed and experienced labour force in both level and percentage - share-of-industry-employment form. Whereas the removal of occupations with zeroes is, no doubt, efficient from a storage cost point of view, it will prove awkward when applying most statistical procedures. That is, since different industries require different combinations of occupations, there will not be a one-to-one correspondence of occupations amongst industries, as comparison techniques like cluster analysis require.

This space-saving file structure, therefore, necessitated a number of file manipulations. The procedure was first to create a blank file of approximately 136,000 records (265 industries multiplied by 511 occupations), which represents the maximum number of records (each identified by the industry and occupation codes) there could have been, given total occupational use in all industries. Then the original combined Census file (described above) was compared to this file. Where there was a match of codes, the 1981 experienced labour force share (which is all we required at this point) was moved to the corresponding slot in the newly-created file, thus leaving blanks where no occupation

existed in the particular industry. The final step was to "transpose" this file containing 265 x 511 records to one containing just 265 records each of dimension 511 (occupations).

The accepted measure of nearness amongst entities in cluster analysis is the correlation coefficient. Therefore, correlations were computed amongst the industries with the occupation as the unit of observation. The resulting correlation matrix was saved on tape/disk, ready for use in cluster analysis (CA). It was felt necessary to do it this way because both space and time would be saved and flexibility gained. First, we would have on file a 265-by-265 matrix instead of one that is 265-by-511. And second, the similarity (correlation) matrix used by CA would not have to be re-calculated each time CA was to be performed on different combinations of industries.

However, when the correlation coefficients were initially calculated on the basis of this full data set, it was discovered that most coefficients were high and close in value. Undoubtedly, this was due to the many zero occupational pairings that exist amongst industries, which likely overwhelm the actual non-zero comparisons and bias the results in their favour. This led us to suspect that factoring techniques like cluster analysis would encounter difficulty in distinguishing amongst industries. And, indeed, as expected when we did apply SAS's Cluster Analysis (CA) to a group of about 50 industries we were left with but one cluster. When we specified as an option a minimum of five clusters we ended up with five; when we set it at ten we got ten. Such results, then, appear to cast doubt on our analysis since this is the very thing we were trying to avoid in the first place: arbitrarily choosing a set number of industries a priori.

Still, the kind of results we did obtain from the above appeared "logical", following fairly closely the lines of the SIC groupings. This would lend credence to the claim that the CA procedure is fairly robust. It thus gave us hope for its eventual success in aggregating industries in an empirically justifiable manner -- if only the problem of the zero coefficients could be solved.

The difficulty is that the closest thing SAS has to what we need is a "pairwise deletion" option in dealing with missing values (taking zeroes as missing values). When comparing two variables, this facility will eliminate an observation from the analysis when it is missing on both or either variable. For our purposes, however, we would like only the former. That is, we would like to remove the incidences (of which there are many in this case) when two industries do not contain the same occupation, but keep the comparisons when the occupation exists in either or both industries. Unfortunately, as we said, SAS can not accomodate us on this point and were forced to send the full 265 x 511 coefficient matrix to I.P. Sharp to be processed by APL. The beauty of APL is that it is extremely flexible in its manipulation of matrices.

The above treatment of zero occupational inputs can be justified in the following manner. It is our view that the Standard Occupation Classification was formulated to encompass the entire spectrum of skills in the Canadian economy but that it was never intended to apply to all industries on an individual basis. For several industries, many occupational categories just do not relate to their job structures. Consequently, we felt confident in eliminating

industry comparisons in which the occupation did not exist in both industries. However, in incidences where an occupation existed in one industry but not the other, we were forced to expand the occupational list so that it would conform in size. That is, to provide a basis of comparison in such cases, we assigned a value of zero (representing insignificant) in the industry where the occupation did not exist and allowed the value that already existed in the other industry.

II. Cluster Analysis: A Brief Explanation

Before proceeding, it might be wise to outline exactly what cluster analysis is and what advantages it offers. Simply put, cluster analysis is a term applied to a family of techniques the aim of which is to group the variables or observations of a given data set into a smaller number of distinct clusters. The components of each cluster should be, by some definition, similar, while each cluster itself is definably different from all other clusters. The purpose, then, of cluster analysis is "to place objects into groups or clusters suggested by the data, not defined a priori, such that the objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar"². Put in other words, the objective is to maximize the similarity within clusters and minimize it among.

Unless used strictly for data summary purposes, the application of cluster analysis should yield clusters that follow more or less along logical lines. They should be "so obvious that they could be picked out, at least in small-scale cases, without the need for complicated mathematical techniques and without a precise definition of the term cluster"³. In fact, the goal is to find "natural" clusters while at the same time reducing the number of variables to more manageable proportions without significant loss of information. In our case, the maximum detail available -- 265 industries -- would be impractical and too fine for meaningful consultation. Furthermore, it would simply not make sense to go out into the field armed

2 SAS Users' Guide: Statistics, 1982 Edition, SAS Institute Inc., Cary, N.C., USA, page 417.

3 Everitt, Brian, Cluster Analysis, 2nd Edition, Heinemann Educational Books, London, UK, 1980, page 7.

with information that indicated that hosiery mills and saw mills should be aggregated. Therefore, the results of our analysis will be tempered with a modicum of rational, qualitative judgment.

In recent years, the popularity of cluster analysis has spread into a number of different areas. Such a multi-disciplinary expansion has led to a proliferation of clustering methods with different definitions of clusters and similarity among objects. For instance, there may be totally disjoint clusters where each object may be placed in one and only one cluster, or overlapping clusters where objects may appear in more than one cluster. Also, the object of the clustering may be either the observation (e.g. occupation) or the variable (e.g. the industry). Finally, one may employ different measures of distance or similarity amongst the entities being analyzed. An example of the latter would be the correlation coefficient while that of the former would be the Euclidian metric.⁴

One advantage of cluster analysis is that its usage is not bogged down in theory, unlike regression analysis whose rigorous tests and procedures often prove too complicated for the layman. Being non-parametric, it can easily be applied and generate results that are readily interpretable and intuitively understood.

4 For explanation of these and other measures, see Everitt, op cit, pp 49-59. The product moment correlation coefficient, as a measure of nearness, has been criticized on the grounds that it places more importance on the similarity of shapes than on actual levels "because the correlation is unity whenever two profiles are parallel, irrespective of how far apart they may be" (Everitt, op cit, p. 53). But, as we indicated earlier, this is the very thing we are interested in: the occupational distribution of the industries, especially considering the data have all been standardized to industry size.

For our purposes we turned to the computer package SAS, the Statistical Analysis System, which maintains a comprehensive set of cluster analysis techniques. Of these procedures, we chose VARCLUS which analyses variables (industries) and generates non-overlapping clusters which is what we required.

Basically, VARCLUS like most cluster analysis procedures, depends upon a clustering criterion which is in fact the formula for the total scatter or dispersion (variation) of the data set. The total is made up of two elements, the within groups variation and the among groups variation. Generally, cluster analysis strives to minimize the former (which is equivalent to maximizing the latter), and assigns variables to clusters based on this criterion. If correlation coefficients are being input, as they are in this case, they may be treated simply as standardized variation measures. To reduce the amount of time and workload, VARCLUS computes and utilizes components which are independent weighted combinations of the variables they represent, and then "tries to maximize the sum across clusters of the variation accounted for by the cluster components".⁵

Looking at the progression of events step-by-step, VARCLUS starts by placing all variables in a single component and then proceeds iteratively through the following stages:

- (i) A cluster is selected for splitting that has the smallest proportion of variation explained by its cluster component;

5 For more details see the relevant chapter in the SAS User's Guide: Statistics, op cit., pp. 462-473, or B. Everitt, op. cit.

- (ii) The chosen cluster is split by finding the first two principal components, rotating them and assigning variables to the cluster with which it is the most correlated;
- (iii) The variables are then reassigned to clusters to maximize the variation explained by the cluster components.

The procedure stops when the criterion is satisfied that stipulates that each cluster has a single eigenvalue greater than one -- implying that only one component of a cluster explains more than the "normal" standardized variation. In other words, the empirically optimal clustering arrangement has been attained at this point.

III. Some Reservations and Issues

The application of cluster analysis to the industry/occupation transformation matrix will provide us with, if nothing else, an empirical justification for the particular configuration of industries in use. To date, we have been using the 69 industry breakdown supplied to us by COFOR and it has never been made clear why that particular structure was chosen over another. Whereas certain data realities almost pre-determine what industry detail can be used in certain sectors, they do not explain the usage in all sectors. At least now we will have empirical corroboration.

Still, there remain reservations with the approach taken, both methodological and conceptual. First, whatever our theoretical criteria tell us in terms of the level of industry disaggregation we should use, it must be remembered that there is only so much detail at which any data source (let alone projection model) can provide us output, employment and other information. Though we would be able to establish functional relationships explaining historical employment in a number of industries, it is unlikely that we could find projected equivalents of several of the variables included. That is, a macro model, like TIM, can only supply projected figures at a fairly aggregative level thus obviating detailed analysis in many cases. Nevertheless, it may be useful, as well as interesting in its own right, to discover empirically what industries differ on the basis of their occupational profiles.

Another point that should be emphasized here concerns the fact that cluster analysis is non-parametric. As we suggested earlier, this is both an advantage and a disadvantage. First, this renders the formal testing of this

technique for statistical significance difficult. However, it still remains a fairly robust and intuitively appealing numerical indication of any association that may exist amongst industries.

Statistical considerations aside, there is also a conceptual difficulty that must be brought to light. That is, theoretical application may yield a particular industrial aggregation that does not fit in with the way employers and unions view the labour market. If our empirically justified breakdown does not conform to the manner in which the market really operates and employer-worker transactions are undertaken, then we have not really gained anything. In fact, this is why we will conduct our analysis the way we intend to, by limiting successive applications of cluster analysis to within certain "natural" groupings of industries.

Another consideration is the fact that industries may not be similar simply on the basis of their occupational profiles but may also display like employment growth rates (despite different population bases). Two industries growing together do not require separate treatment because a weighted average of the two occupational profiles would provide us with an equivalent amount of information at less cost. A specific illustration of this is the case of vertically integrated industries where one industry depends crucially on the other for its existence (and vice-versa).⁶ Simple

⁶ In point of fact, we mean implicitly integrated industries where, for example, the forestry industry depends on wood products and pulp and paper industries for much of its sales. In this case, output and hence employment in one industry should grow in lock step with the other.

application of cluster analysis will, of course, not capture these effects and this is something we must consider after receiving the initial results.

Yet another point is that it is likely that what would be a sound industrial structure at one time may not be as relevant in another on account of different economic and technological conditions. In fact, it might be interesting to conduct cluster analysis also upon the similarly coded 1971 coefficients. By comparing the resulting industry clusters that emerge to the 1981 clusters, we would be able to detect changing industrial structures reflecting different economic and technological states.

One final issue that must be raised here is that of size. Whatever industry groupings are obtained they still must be placed in perspective. In other words, we must ask the question: "Are each of the clusters large enough to warrant our separate treatment?" An insignificant industry in terms of labour force share may not be worth our while to project employment in, despite the fact that its occupational structure is unique. Instead, we could, if there were no objections from the parties concerned, simply include this industry in the cluster with which it is most strongly associated. Therefore, we will be investigating the relative size of the resulting clusters as well.

All these points will be addressed in turn after completion of the cluster analyses. They are raised here merely for discussion purposes and to emphasize the fact that similarity of occupational structures is not the sole criterion for grouping industries.

IV. Findings

a) Preliminary Results

Reservations aside, preliminary analysis on the APL-processed data set of correlation coefficients has revealed some interesting results. For one thing, the application of cluster analysis to twenty-nine disparate sectors has yielded clusters that confirm many of our pre-conceived notions about which groups they would be associated with. In fact, the clusters approximated the SIC groupings. This is interesting because, based purely on (the differences in) occupational profiles, industries appear to follow a classification system (SIC) that is based on product, input as well as production process.⁷ Indeed it would be interesting to apply cluster analysis to all 265 industries at one time to observe how closely the SIC is matched. In this way we would be entering the exercise with a, in a manner of speaking, blank slate -- without confining the analysis to certain logical divisions and without any a priori biases.

As intellectually rewarding as this may be, however, the application of cluster analysis to all 265 industries may not be practical. For one thing it would severely tax the core requirements of SAS. For another it may produce, as we had alluded to earlier, clusters that lack a firm footing and are rationally indefensible. In other words, the clusters, though statistically sound, may be logically feeble. For these reasons we will be forced to apply cluster analysis on a piece-meal basis as data availability dictates, that is within SIC Divisions.

⁷ For a broader description of the criteria the SIC uses to delineate the industries, see Dominion Bureau of Statistics, Standard Industrial Classification Manual, Cat. no. 12-501, Information Canada, Ottawa, Dec. 1970, pp. 11, 12.

b) Final Results of Cluster Analysis

In this section, the results of applying cluster analysis to the occupational profiles of the 265 census industries are presented by broad industry categories. The actual cluster groupings are reported in the table. Before proceeding to a brief discussion of them, it should be borne in mind (as we have indicated earlier) that these findings are by no means the final word on industry disaggregation. Certainly, the relative size of the resulting clusters, data constraints, the fact that some industries are closely tied with others, and consultative needs will all undoubtedly alter this empirically determined industrial structure. Although our primary concern is occupations, it is by no means the only factor in the decision to treat industries separately or not.

Before describing the sectors on an individual basis, a look at one aspect of that table will reveal one of the major advantages of cluster analysis. Therein is reported the variation explained in percentage terms. The underlined figures opposite the division (sector) title tell us the proportion of total variation in all industries in the corresponding sector explained by the calculated components; the figures below the underlined one indicate the proportion of variation within each individual cluster explained by the corresponding cluster component. As observed there, most shares are greater than one-half, indicating that the majority of variation has been explained by the calculated instrument. The clusters with less than a half normally represent ones, described later, that contain unlikely industries with apparently little in common. In this way, we

INDUSTRY GROUPINGS ACCORDING TO CLUSTER ANALYSIS, VARIATION EXPLAINED AND LABOUR FORCE SHARES

Cluster No.	Cluster Description	Industries Clustered ¹ (SIC codes)	Variation ² Explained (percent)	Labour Force Share ³ 1971 (per cent)	Labour Force Share ³ 1981 (per cent)
<u>DIV. 1: AGRICULTURE</u>					
1	Farms	001-019	100.	5.95	3.99
2	Agricultural Services	021	100.	0.12	0.16
<u>DIV. 2: FORESTRY</u>					
3	Logging	031	100.	0.80	0.67
4	Forestry Services	039	100.	0.14	0.19
<u>DIV. 3: FISHING, TRAPPING, HUNTING</u>					
5	Fishing	041	100.	0.29	0.27
6	Fishing Services	045	100.	0.02	0.04
7	Hunting, Trapping	047	100.	0.02	0.01
<u>DIV. 4: MINING</u>					
8	Metal, Coal, Peat, Misc. Non-Metal	051-061, 072, 079	78.	1.06	0.87
9	Petroleum, Natural Gas, Drilling	064, 096, 099	76.	0.45	0.76
10	Non-metal	071, 073, 083, 087	76.	0.19	0.15
11	Other Contract Drilling	098	100.	0.04	0.03
<u>DIV. 5: MANUFACTURING</u>					
12	Metal Fabr, Agr Impl, Misc Mchry, Railroad Stock	301-304, 307, 309-315, 326, 329	70.	2.05	2.01
13	Glove, Luggage, Textile Products, Clothing	175, 179, 187-249	67.	1.81	1.66
14	Chemicals (less ind. chemicals)	374-377, 379	80.	0.61	0.49
15	Lamp, Fridge, Applces, Light Fix., Elec. Equip.	268, 316, 331-333, 336-339	78.	0.87	0.26
16	Textiles, Jewellery	181-186, 392	77.	0.63	0.45
17	Leather, Clay, Abrasives, Lime	172, 351, 357-359	57.	0.26	0.22
18	Primary Metals	291-298	48.	1.47	1.19
19	Food less Fish, Bakery, Misc.	101, 103-106, 109	73.	1.71	1.33
20	Sash, Door, Caskets, Furniture	254, 258, 261-266	46.	0.73	0.86
21	Printing, Publishing	286-289	85.	1.30	1.18

22	Sporting Goods, Signs, Misc.	393-399	73.	0.45	0.41
23	Office Mchry, Radio/TV, Comm Eqpt, Scientific Eqpt	318, 334, 335, 391	59.	1.11	0.91
24	Cement, Concrete, Misc. Petrol Prds	352, 354, 355, 369	67.	0.30	0.26
25	Paper and Allied less Asphalt	271, 273, 274	66.	1.55	1.25
26	Wire, Tool, Machines, Aircraft	305, 306, 308, 321	54.	0.94	0.94
27	Ship and Boat Building	327, 328	89.	0.23	0.21
28	Wood Box, Misc. Wood	256, 259	89.	0.12	0.13
29	Rubber, Plastics	162, 165	85.	0.56	0.60
30	Motor Vehicles, Truck Body	323, 324	88.	1.30	1.11
31	Asphalt, Petroleum Products, Fertilizers, etc.	172, 365, 372, 373, 378	63.	0.63	0.60
32	Saw and Veneer Mills	251, 252	77.	0.94	0.88
33	Tobacco	151, 153	92.	0.11	0.07
34	Stone, Glass Products	353, 356	71.	0.19	0.15
35	Fish Products, Shoe Factories	102, 174	53.	0.61	0.70
36	Bakery, Misc. Food	107, 108	79.	0.96	0.78

DIV. 6: CONSTRUCTION

37	Building, Special Trade	404, 421	84.	5.35	5.39
38	Highway, Other General Construction	406, 409	85.	1.43	1.10

DIV. 7: TRANSPORTATION, STORAGE, COMMUNICATION, UTILITIES

39	Moving, Truck, Hwy. Maintenance, Other utilities	506, 507, 516, 579	70.	1.87	1.73
40	Bus, Other Transportation	508, 509, 519	98.	0.50	0.61
41	Pipeline, Utilities	515, 572-576	62.	1.07	1.07
42	Storage	524, 527	68.	0.20	0.19
43	Air Transport, Misc. Transportation Services	501, 502, 517	61.	0.57	0.80
44	Communication less Post Office	543-545	63.	1.34	1.47
45	Rail, Water	503-505	49.	1.87	1.27
46	Taxicab Operations	512	100.	0.31	0.25
47	Post Office	548	100.	0.71	0.67

DIV. 8: TRADE

48	Retail, non-auto related	631-652, 663-676, 681-694, 696-699	79.	8.92	9.53
49	Retail, automotive	654-658	82.	2.53	2.42
50	Small Appliance Repair	678	100.	0.11	0.09
51	Watch, Jewellery Repair	695	100.	0.02	0.02
52	Wholesale to primary-related inds.	602, 606, 608, 622, 626, 627	71.	1.25	1.25

53	Wholesale to non-primary	611-621, 623-625, 629	60.	3.13	3.57
<u>DIV. 9: FINANCE, INSURANCE, REAL ESTATE</u>					
54	Finance	701-707	65.	2.33	2.74
55	Insurance, Real Estate	721-737	65.	2.18	2.61
<u>DIV. 10: SERVICES</u>					
56	Employment Agencies, Lawyers, Labour Org.	851, 866, 891, 899	71.	1.23	1.85
57	Hospitality, Building Services	881-886, 898	59.	4.59	6.26
58	Health less Doctors' Offices	821, 822, 824, 826, 827	54.	4.92	5.06
59	Elementary, Secondary Education	801, 802	60.	5.21	4.54
60	Rental, Misc. Repair	894, 895, 897	88.	0.26	0.35
61	Film Processing - Movies, Photos	842, 893	61.	0.16	0.23
62	Post-Secondary, non-univ.	804, 805, 809	63.	0.43	0.56
63	Arts Schools, Blacksmith/Welding	803, 896	83.	0.19	0.25
64	Architects', Engineers' Offices	863, 864	52.	0.71	1.02
65	Welfare, Household, Misc. Pers. Serv.	828, 873, 879	70.	1.72	1.96
66	Shoe Repair	871	63.	0.04	0.02
67	Funeral	877	100.	0.09	0.07
68	Accountants, Bus. Consultants, Misc. Bus. Serv.	861, 867, 869	100.	0.87	1.45
69	Security	855	67.	0.21	0.31
70	Religious	831	100.	0.59	0.53
71	Dentists' Offices, Theatrical	825, 845	53.	0.30	0.47
72	Movie Houses, Bowling Alleys, Golf Clubs	841, 843, 844, 849	58.	0.78	0.94
73	Universities/Colleges	806	100.	1.24	1.19
74	Advertising Offices	862	100.	0.18	0.22
75	Barbers, Hairdressers	872	100.	0.69	0.56
76	Computer	853	100.	0.06	0.27
77	Laundry	874, 876	87.	0.51	0.35
78	Physicians' Offices	823	100.	0.55	0.59
79	Libraries, Museums	807	100.	0.18	0.23
<u>DIV. 11: PUBLIC ADMINISTRATION</u>					
80	Defence	902	88.	1.76	1.06
81	Federal, Provincial	909, 931, 991	100.	4.45	4.53
82	Local	951	80.	1.85	2.05
			100.		

Notes

- 1 For explanation of codes, see Appendix.
- 2 Figure underlined represents the proportion of total variation in the sector industries explained by the cluster components; the figures below represent the variation in the individual clusters explained by the component.
- 3 Figures represent the proportion of total experienced labour force possessed by the respective clusters according to the 1971 and 1981 Censuses, where experienced is defined as having some labour market experience during the 18 months prior to June of the Census year.

have a measure -- cluster analysis -- that not only gives us intuitively comprehensible results but also provides a quantitative feel for the explanatory power of the clusters.

Also on that table are listed the 3-digit SIC codes of the industries which make up each cluster. The definitions of these codes is provided in an Appendix. In addition, there are the relative labour force shares, in 1971 and 1981, of those groupings. Such information will not be discussed here but is presented for discussion purposes only.

Agriculture

Applying cluster analysis to the occupational profiles of the two agricultural industries on the census tape has revealed that the two sectors -- farms and services incidental to agriculture -- are significantly different. This is a change from our previous 69 industry disaggregation which simply divided up agriculture into paid and other-than-paid categories.

Forestry

Again we see cluster analysis has deemed that a sector -- forestry -- should be split into separate sub-sectors -- logging and forestry services -- where formerly the two sub-groups were lumped into one and re-divided into paid and other-than-paid. It would seem that the occupational skills required by logging and those that are needed in forestry services, which include patrolling, fire fighting, nurseries and reforestation, are quite a bit different.

Fishing, Hunting and Trapping

In this sector, by reason of differing occupational profiles, there are now three sub-sectors where formerly there was one. It would seem logical that the occupations required in fishing would diverge from those required by fishery services which includes hatcheries, inspection and protection and the skills needed by hunters and trappers. Still, it is unlikely that in light of data availability and the relative size of the industries concerned, these sectors will be kept as separate units.

Mining

In the mining sector, cluster analysis based on occupational profiles has grouped industries that in some cases appear to be unlike. However it may be, and it was proven upon inspection of the actual occupational distributions, that the industries do have similar occupational input requirements. Questions could certainly, at least initially, have been raised concerning the lumping of coal, peat and miscellaneous non-metal mining with the metal mining industries of gold, uranium and iron.

Manufacturing

Along with the service industries, manufacturing proved to be one of the more difficult sectors to satisfactorily apply cluster analysis to, in the sense of achieving logical, explicable results. The problem undoubtedly lies in the fact that there is a myriad of dissimilar industries -- with a corresponding multitude of different occupational profiles -- all included under the umbrella of manufacturing.

The fact that there are several unique distributions within manufacturing makes it difficult for the cluster analysis procedure to decide which sectors should be associated with which cluster, leading to some apparently anomolous groupings as table 1 would indicate. Take for example cluster 16 which groups textiles with jewellery, cluster 26 which brings together aircraft with wire, tools and machines and cluster 35 which lumps fish products with shoes.⁸

At the same time, there are a number of groupings which, though not parallel to the SIC in the sense of being in contiguous order, do appear logical upon second thought. Illustrative of this phenomenon is cluster 15 which contains lamps, refridgerators, appliances, light fixtures and electrical wire and cluster 23 which groups office machinery, radio and televisions, communication equipment and scientific instruments. Regarding the former cluster, it is interesting to note that cluster analysis has removed electrical lamps from the furniture and fixtures categories and placed it with light fixtures. Also, observe that the procedure has singled out shipbuilding and boat building to stand alone which may have some implication for sector studies.

In passing it should be noted that the above groupings are quite a bit different from the manufacturing breakdown contained in the original 69-industry disaggregation which stuck mainly to the 2-digit SIC. However, as we demonstrated, the SIC may not necessarily categorize industries in the most realistic manner.

⁸ It should be pointed out that an industry which conceptually does not appear to belong to the cluster in which it has been placed normally displayed a low degree of adhesion (correlation) to its own cluster and/or a high correlation to the next closest cluster. Such measures are not listed here but are available on request.

Construction

This is another problematic sector, although for different reasons. As will be fleshed out in the Construction Sector Study Proposal, this sector covers a host of different construction types each requiring different combinations of skill inputs. Unfortunately the Census has broken down this industry into only four sub-sectors, surely not exhaustive enough to accurately cover this area of the economy. Still, on the basis of cluster analysis of the industries' occupational distributions special trade contractors (code 421) have been grouped with building construction (code 404). According to the SIC manual (1970)⁹, special trade contractors are primarily "engaged in repair and maintenance work, done at site, on buildings of all types", making this pairing quite logical. Moreover, the fact that highway, street and bridge construction (code 406) is combined with other general construction (code 409), which includes waterworks, gas mains, and railway right-of-ways, seems to be fairly logical. Indeed, looking at the construction sector in this manner makes more sense than the paid/unpaid breakdown of the previous list.

Transportation, Storage, Communication, Utilities

The application of cluster analysis to this sector provided some interesting results. For one thing, CA has grouped moving and storage, other truck transport, highway and bridge maintenance and other utilities (which includes garbage disposal) into one cluster, a reasonable conclusion given that all these industries are, to a certain extent, dependent on the truck. Another interesting aggregation is that of bus transport with other transport which includes sightseeing and school buses. Even more fascinating is the

⁹ See Dominion Bureau of Statistics, op. cit., p. 33.

inclusion of pipelines with utilities (less 'other' utilities), especially since pipeline transport is mainly involved with distribution of commodities through pipelines. Also of note is the isolation of taxicab operations and the post office, the latter being removed from the other more "traditional" communication industries like telephone and radio and television. Storage is kept exactly as the SIC has categorized this industry.

Two oddities do emerge, however. One is the grouping of air transportation and miscellaneous transportation services, unless, of the lines of business the latter contain, tourist/ticket agencies make up the majority. Equally unusual is the combining of rail and water transportation.

Trade

Trade turned out to be another fruitful area for cluster analysis. For, it formed six clusters, where formerly there were but two in the original 69 (paid/unpaid) and the 1970 SIC (retail/wholesale), which seems a much more relevant breakdown of trade. Such a feeling in fact is revealed in the 1980 SIC¹⁰ which now designates 16 categories of trade. Still, cluster analysis has retained the retail/wholesale character of this sector by keeping each cluster wholly within one or the other major group. For instance, CA has split the retail side four ways: automotive-related retailers (except tires) in one cluster, non-automotive related (like food, clothing) in another, and small appliance repair and watch/jewellery repair as their own separate clusters. The wholesale trade disaggregation

¹⁰ Statistics Canada, Standard Industrial Classification 1980, Supply and Services Canada, December 1980.

appears to be along primary/non-primary lines with cluster 52 containing businesses selling products pertaining to primary industries (eg. farm products, coal and coke, and lumber) and cluster 53 containing those that do not.

Finance, Insurance, Real Estate

This sector was fairly reasonably split by cluster analysis. As is revealed in the table, finance-related industries like banks and security brokers are grouped in one cluster (like the SIC) while insurance and real estate firms are in another. It is interesting to note that CA has associated insurance carriers (code 721) with insurance and real estate agencies (code 735) and real estate operators (code 737) as one would expect but not as the SIC has placed them.

Services

This sector presents another interesting challenge for cluster analysis, mainly because it contains a multitude of heterogeneous industries many with unique occupational profiles. This is directly manifested in the appearance of several clusters with but one industry. Indeed, of the 24 clusters formed, 10 are single-industry ones, most of which should, intuitively, stand out on their own. It would be hard to decide with which, if any, industries one should group shoe repair, funeral, security, religious, barbering and laundry services.

As for the multi-industry clusters, most appear to be "logical". For instance, the fact that building services

like janitorial and cleaning are combined with hospitality industries like hotels, restaurants and taverns may be explained by the fact that many of the latter establishments hire their own staff to provide the former services. Also, the fact two non-contiguous SIC industries -- motion pictures production (code 842) and photography and finishing (code 893) -- appear together in cluster 61 seems reasonable enough given that both are involved essentially in the same commodity, film: one still and one moving. It is also interesting to observe that CA has logically removed doctors' offices (codes 823 and 825) from the other health services (like the original 69) and that it has trichotomized educational services into "non-post-secondary" (cluster 59), non-university/college post-secondary (cluster 62) and university and college (cluster 73).

Finally two unusual associations have been obtained: cluster 63 which pairs schools of art (803) with blacksmithing/welding (896) and cluster 71, dentists' offices and theatrical services.

Public Administration

Cluster analysis has split this sector three ways. Diverging somewhat from the SIC, defence (code 902) has been removed from federal administration to stand on its own, while federal and provincial governments have been grouped -- logically enough -- in cluster 81. Finally, the uniqueness of local administration dictates that this level of administration should stand alone. Such a breakdown should indicate that the fact that the original 69 left this sector undivided may not have been the correct approach.

Conclusion

In sum, we may observe that cluster analysis (CA) has provided us with an empirically verifiable industry breakdown, containing some 80 industries out of the original 265 industries, based purely on differences in the occupational profiles. In addition, CA has provided us with a device to reduce the amount of information in both a methodologically sound and, in most cases, logically defensible manner. The fact that many of the resulting clusters cut across SIC major groups may not be a problem since, in most instances, it was a reasonable association and perhaps reflective of true labour market conditions and the way most people view the industrial world. This, combined with the fact that one can measure the amount of variation explained by the clusters, as well as the degree of association among and within clusters, points to the robustness of this technique.

However, as we have stated several times before, although cluster analysis has introduced some interesting ideas, that would only before have been suspected and not empirically demonstrated, this procedure is only a minimum requirement for a workable industry disaggregation. As we have mentioned, cluster significance, in terms of share of total labour force, has to enter the decision as to the final breakdown. In addition, similarity of employment growth rates and the degree of implicit integration amongst industries will also determine their separability. Finally, operability, both from a data availability and consultative standpoint, will, as always, be the last word on what our final list of industries will look like.

In sum, then, we have used a novel, yet fairly rigorous, approach to obtain a set of disaggregated industries. Our primary aims were to empirically arrive at an effective industrial breakdown while at the same time documenting our work towards that end. Those goals we have, to a certain extent, met. The question now to be asked is: How reasonable is it and will it be useful in both our estimation exercise and our private sector dealings? The attempt to answer these questions will comprise the next stage of our analysis.

APPENDIX

STANDARD INDUSTRIAL CLASSIFICATION (1970)

Division 1—Agriculture		Ind.	Division 4—Mines (including Milling), Quarries
		No.	and Oil Wells — Concluded
Major Group 1—Experimental and Institutional Farms			Major Group 5—Services Incidental to Mining
001	Experimental and University Farms	096	Contract Drilling for Petroleum
003	Institutional Farms	098	Other Contract Drilling
		099	Miscellaneous Services Incidental to Mining
Major Group 2—Farms (except Experimental and Institutional Farms)			Division 5—Manufacturing Industries
011	Livestock and Livestock Combination Farms		Major Group 1—Food and Beverage Industries
013	Field Crop and Field Crop Combination Farms	101	Meat and Poultry Products Industries
015	Fruit and Vegetable Farms	102	Fish Products Industry
017	Other Crop and Livestock Combination Farms	103	Fruit and Vegetable Processing Industries
019	Miscellaneous Specialty Farms	104	Dairy Products Industry
		105	Flour and Breakfast Cereal Products Industry
Major Group 3—Services Incidental to Agriculture		106	Feed Industry
021	Services Incidental to Agriculture	107	Bakery Products Industries
		108	Miscellaneous Food Industries
		109	Beverage Industries
Division 2—Forestry			Major Group 2—Tobacco Products Industries
Major Group 1—Logging		151	Leaf Tobacco Processors
031	Logging	153	Tobacco Products Manufacturers
Major Group 2—Forestry Services			Major Group 3—Rubber and Plastics Products Industries
039	Forestry Services	162	Rubber Products Industries
		165	Plastics Fabricating Industry, n.e.s.
Division 3—Fishing and Trapping			Major Group 4—Leather Industries
Major Group 1—Fishing		172	Leather Tanneries
041	Fishing	174	Shoe Factories
		175	Leather Glove Factories
Major Group 2—Fishery Services		179	Luggage, Handbag and Small Leather Goods Manufacturers
045	Fishery Services		Major Group 5—Textile Industries
Major Group 3—Hunting and Trapping		181	Cotton Yarn and Cloth Mills
047	Hunting and Trapping	182	Wool Yarn and Cloth Mills
		183	Man-made Fibre, Yarn and Cloth Mills
Division 4—Mines (including Milling), Quarries and Oil Wells		184	Cordage and Twine Industry
Major Group 1—Metal Mines		185	Felt and Fibre Processing Mills
051	Placer Gold Mines	186	Carpet, Mat and Rug Industry
052	Gold Quartz Mines	187	Canvas Products, and Cotton and Jute Bags Industries
057	Uranium Mines	188	Automobile Fabric Accessories Industry
058	Iron Mines	189	Miscellaneous Textile Industries
059	Miscellaneous Metal Mines		Major Group 6—Knitting Mills
Major Group 2—Mineral Fuels		231	Hosiery Mills
061	Coal Mines	239	Knitting Mills (except Hosiery Mills)
064	Crude Petroleum and Natural Gas Industry		Major Group 7—Clothing Industries
Major Group 3—Non-Metal Mines (except Coal Mines)		243	Men's Clothing Industries
071	Asbestos Mines	244	Women's Clothing Industries
072	Peat Extraction	245	Children's Clothing Industry
073	Gypsum Mines	246	Fur Goods Industry
079	Miscellaneous Non-Metal Mines	248	Foundation Garment Industry
		249	Miscellaneous Clothing Industries
Major Group 4—Quarries and Sand Pits			Major Group 8—Wood Industries
083	Stone Quarries	251	Sawmills, Planing Mills and Shingle Mills
087	Sand Pits or Quarries	252	Veneer and Plywood Mills
		254	Sash, Door and Other Millwork Plants
		256	Wooden Box Factories
		258	Coffin and Casket Industry
		259	Miscellaneous Wood Industries

Division 5—Manufacturing Industries — Continued**Major Group 9—Furniture and Fixture Industries**

- 261 Household Furniture Manufacturers
- 264 Office Furniture Manufacturers
- 266 Miscellaneous Furniture and Fixtures Manufacturers
- 268 Electric Lamp and Shade Manufacturers

Major Group 10—Paper and Allied Industries

- 271 Pulp and Paper Mills
- 272 Asphalt Roofing Manufacturers
- 273 Paper Box and Bag Manufacturers
- 274 Miscellaneous Paper Converters

Major Group 11—Printing, Publishing and Allied Industries

- 286 Commercial Printing
- 287 Platemaking, Typesetting and Trade Bindery Industry
- 288 Publishing Only
- 289 Publishing and Printing

Major Group 12—Primary Metal Industries

- 291 Iron and Steel Mills
- 292 Steel Pipe and Tube Mills
- 294 Iron Foundries
- 295 Smelting and Refining
- 296 Aluminum Rolling, Casting and Extruding
- 297 Copper and Copper Alloy Rolling, Casting and Extruding
- 298 Metal Rolling, Casting and Extruding, n.e.s.

Major Group 13—Metal Fabricating Industries (except Machinery and Transportation Equipment Industries)

- 301 Boiler and Plate Works
- 302 Fabricated Structural Metal Industry
- 303 Ornamental and Architectural Metal Industry
- 304 Metal Stamping, Pressing and Coating Industry
- 305 Wire and Wire Products Manufacturers
- 306 Hardware, Tool and Cutlery Manufacturers
- 307 Heating Equipment Manufacturers
- 308 Machine Shops
- 309 Miscellaneous Metal Fabricating Industries

Major Group 14—Machinery Industries (except Electrical Machinery)

- 311 Agricultural Implement Industry
- 315 Miscellaneous Machinery and Equipment Manufacturers
- 316 Commercial Refrigeration and Air Conditioning Equipment Manufacturers
- 318 Office and Store Machinery Manufacturers

Major Group 15—Transportation Equipment Industries

- 321 Aircraft and Aircraft Parts Manufacturers
- 323 Motor Vehicle Manufacturers
- 324 Truck Body and Trailer Manufacturers
- 325 Motor Vehicle Parts and Accessories Manufacturers
- 326 Railroad Rolling Stock Industry
- 327 Shipbuilding and Repair
- 328 Boatbuilding and Repair
- 329 Miscellaneous Vehicle Manufacturers

Major Group 16—Electrical Products Industries

- 331 Manufacturers of Small Electrical Appliances
- 332 Manufacturers of Major Appliances (Electric and Non-Electric)
- 333 Manufacturers of Lighting Fixtures
- 334 Manufacturers of Household Radio and Television Receivers
- 335 Communications Equipment Manufacturers
- 336 Manufacturers of Electrical Industrial Equipment
- 338 Manufacturers of Electric Wire and Cable
- 339 Manufacturers of Miscellaneous Electrical Products

Division 5—Manufacturing Industries — Concluded**Major Group 17—Non-Metallic Mineral Products Industries**

- 351 Clay Products Manufacturers
- 352 Cement Manufacturers
- 353 Stone Products Manufacturers
- 354 Concrete Products Manufacturers
- 355 Ready-Mix Concrete Manufacturers
- 356 Glass and Glass Products Manufacturers
- 357 Abrasives Manufacturers
- 358 Lime Manufacturers
- 359 Miscellaneous Non-Metallic Mineral Products Industries

Major Group 18—Petroleum and Coal Products Industries

- 365 Petroleum Refineries
- 369 Miscellaneous Petroleum and Coal Products Industries

Major Group 19—Chemical and Chemical Products Industries

- 372 Manufacturers of Mixed Fertilizers
- 373 Manufacturers of Plastics and Synthetic Resins
- 374 Manufacturers of Pharmaceuticals and Medicines
- 375 Paint and Varnish Manufacturers
- 376 Manufacturers of Soap and Cleaning Compounds
- 377 Manufacturers of Toilet Preparations
- 378 Manufacturers of Industrial Chemicals
- 379 Miscellaneous Chemical Industries

Major Group 20—Miscellaneous Manufacturing Industries

- 391 Scientific and Professional Equipment Industries
- 392 Jewellery and Silverware Industry
- 393 Sporting Goods and Toy Industries
- 397 Signs and Displays Industry
- 399 Miscellaneous Manufacturing Industries, n.e.s.

Division 6—Construction Industry**Major Group 1—General Contractors**

- 404 Building Construction
- 406 Highway, Bridge and Street Construction
- 409 Other Construction

Major Group 2—Special-Trade Contractors

- 421 Special-Trade Contractors

Division 7—Transportation, Communication and Other Utilities**Major Group 1—Transportation**

- 501 Air Transport
- 502 Services Incidental to Air Transport
- 503 Railway Transport
- 504 Water Transport
- 505 Services Incidental to Water Transport
- 506 Moving and Storage, Used Goods, Uncrated
- 507 Other Truck Transport
- 508 Bus Transport, Interurban and Rural
- 509 Urban Transit Systems
- 512 Taxicab Operations
- 515 Pipeline Transport
- 516 Highway and Bridge Maintenance
- 517 Miscellaneous Services Incidental to Transport
- 519 Other Transportation

Major Group 2—Storage

- 524 Grain Elevators
- 527 Other Storage and Warehousing

**Division 7 – Transportation, Communication
and Other Utilities – Concluded**

Major Group 3 – Communication

- 543 Radio and Television Broadcasting
- 544 Telephone Systems
- 545 Telegraph and Cable Systems
- 548 Post Office

Major Group 4 – Electric Power, Gas and Water Utilities

- 572 Electric Power
- 574 Gas Distribution
- 576 Water Systems
- 579 Other Utilities

Division 8 – Trade

Major Group 1 – Wholesale Trade

- 602 Wholesalers of Farm Products
- 606 Wholesalers of Coal and Coke
- 608 Wholesalers of Petroleum Products
- 611 Wholesalers of Paper and Paper Products
- 612 Wholesalers of General Merchandise
- 614 Wholesalers of Food
- 615 Wholesalers of Tobacco Products
- 616 Wholesalers of Drugs and Toilet Preparations
- 617 Wholesalers of Apparel and Dry Goods
- 618 Wholesalers of Household Furniture and Furnishings
- 619 Wholesalers of Motor Vehicles and Accessories
- 621 Wholesalers of Electrical Machinery, Equipment and Supplies
- 622 Wholesalers of Farm Machinery and Equipment
- 623 Wholesalers of Machinery and Equipment, n.e.s.
- 624 Wholesalers of Hardware, Plumbing and Heating Equipment
- 625 Wholesalers of Metal and Metal Products, n.e.s.
- 626 Wholesalers of Lumber and Building Materials
- 627 Wholesalers of Scrap and Waste Materials
- 629 Wholesalers, n.e.s.

Major Group 2 – Retail Trade

- 631 Food Stores
- 642 General Merchandise Stores
- 652 Tire, Battery and Accessories Stores
- 654 Gasoline Service Stations
- 656 Motor Vehicle Dealers
- 658 Motor Vehicle Repair Shops
- 663 Shoe Stores
- 665 Men's Clothing Stores
- 667 Women's Clothing Stores
- 669 Clothing and Dry Goods Stores, n.e.s.
- 673 Hardware Stores
- 676 Household Furniture and Appliance Stores
- 678 Radio, Television and Electrical Appliance Repair Shops
- 681 Drug Stores
- 691 Book and Stationery Stores
- 692 Florists' Shops
- 694 Jewellery Stores
- 695 Watch and Jewellery Repair Shops
- 696 Liquor, Wine and Beer Stores
- 697 Tobacconists
- 699 Retail Stores, n.e.s.

Division 9 – Finance, Insurance and Real Estate

Major Group 1 – Finance Industries

- 701 Banks and Other Deposit Accepting Establishments
- 703 Other Credit Agencies
- 705 Security Brokers and Dealers (including Exchanges)
- 707 Investment and Holding Companies
- 715 Canadian Offices of Canadian-Incorporated Companies Classified as Non-Canadian

**Division 9 – Finance, Insurance
and Real Estate – Concluded**

Major Group 2 – Insurance Carriers

- 721 Insurance Carriers

Major Group 3 – Insurance Agencies and Real Estate Industry

- 735 Insurance and Real Estate Agencies
- 737 Real Estate Operators

Division 10 – Community, Business and Personal Service Industries

Major Group 1 – Education and Related Services

- 801 Kindergartens and Nursery Schools
- 802 Elementary and Secondary Schools
- 803 Schools of Art and of the Performing Arts
- 804 Vocational Centers, Trade Schools and Business Colleges
- 805 Post-Secondary Non-University Educational Institutions
- 806 Universities and Colleges
- 807 Libraries, Museums and Other Repositories
- 809 Education and Related Services, n.e.s.

Major Group 2 – Health and Welfare Services

- 821 Hospitals
- 822 Related Health Care Institutions
- 823 Offices of Physicians and Surgeons
- 824 Offices of Para-medical Personnel (Practitioners)
- 825 Offices of Dentists
- 826 Diagnostic and Therapeutic Services, n.e.s.
- 827 Miscellaneous Health Services
- 828 Welfare Organizations

Major Group 3 – Religious Organizations

- 831 Religious Organizations

Major Group 4 – Amusement and Recreation Services

- 841 Motion Picture Theatres
- 842 Motion Picture Production and Distribution
- 843 Bowling Alleys and Billiard Parlours
- 844 Golf Clubs and Country Clubs
- 845 Theatrical and Other Staged Entertainment Services
- 849 Miscellaneous Amusement and Recreation Services

Major Group 5 – Services to Business Management

- 851 Employment Agencies and Personnel Suppliers
- 853 Computer Services
- 855 Security and Investigation Services
- 861 Offices of Accountants
- 862 Advertising Services
- 863 Offices of Architects
- 864 Engineering and Scientific Services
- 866 Offices of Lawyers and Notaries
- 867 Offices of Management and Business Consultants
- 869 Miscellaneous Services to Business Management

Major Group 6 – Personal Services

- 871 Shoe Repair Shops
- 872 Barber and Beauty Shops
- 873 Private Households
- 874 Laundries, Cleaners and Pressers (except Self-Service)
- 876 Self-service Laundries and Dry Cleaners
- 877 Funeral Services
- 879 Miscellaneous Personal Services

Major Group 7 – Accommodation and Food Services

- 881 Hotels and Motels
- 883 Lodging Houses and Residential Clubs
- 884 Camping Grounds and Trailer Parks
- 886 Restaurants, Caterers and Taverns

**Division 10 – Community, Business and Personal Service
Industries – Concluded**

Major Group 8 – Miscellaneous Services

891 Labour Organizations and Trade Associations
893 Photographic Services, n.e.s.
894 Automobile and Truck Rental
895 Machinery and Equipment Rental
896 Blacksmithing and Welding Shops
897 Miscellaneous Repair Shops
898 Services to Buildings and Dwellings
899 Miscellaneous Services, n.e.s.

Division 11 – Public Administration and Defence

Major Group 1 – Federal Administration

902 Defence Services
909 Other Federal Administration

**Division 11 – Public Administration
and Defence – Concluded**

Major Group 2 – Provincial Administration

931 Provincial Administration

Major Group 3 – Local Administration

951 Local Administration

Major Group 4 – Other Government Offices

991 Other Government Offices

Division 12 – Industry Unspecified or Undefined

999 Canadian-Incorporated Companies Classified as Non-
Canadian
000 Unspecified or Undefined

Source: Dominion Bureau of Statistics, Standard Industrial Classification
Manual, Revised 1970, Information Canada, Ottawa, December, 1970.

Source: Bureau fédéral de la statistique, Classification des activités économiques, révisée 1970, Information Canada, Ottawa, juin 1971.

884	Terrains de camping et parcs à roulettes	Groupe 7 - Hébergement et restauration - fin	884	Divisions 10 - Services socio-culturels, commerciaux et personnels - fin	884	Divisions 11 - Administration publique et défense - fin
885	Restaurants, traiteurs et lavernes		885	Administration provinciale	885	Administration provinciale
891	Groupe 8 - Services divers		891	Groupe 3 - Administration locale	891	Groupe 3 - Administration locale
892	Syndicats ouvriers et associations professionnelles		892	Administration locale	892	Administration locale
893	Photographie, n.c.a.		893	Groupe 4 - Bureaux de gouvernements étrangers	893	Groupe 4 - Bureaux de gouvernements étrangers
894	Location d'automobiles et de camions		894	Bureaux de gouvernements étrangers	894	Bureaux de gouvernements étrangers
895	Location de machines et de matériel		895	Ateliers de réparations diverses	895	Ateliers de réparations diverses
896	Maréchaussée et soudure		896	Entretien de bâtiments et d'habitations	896	Entretien de bâtiments et d'habitations
897	Services divers, n.c.a.		897	Services divers, n.c.a.	897	Services divers, n.c.a.
902	Divisions 11 - Administration publique et défense		902	Divisions 12 - Activités indéterminées et imprécises	902	Divisions 12 - Activités indéterminées et imprécises
909	Groupe 1 - Administration fédérale		909	Sociétés constituées au Canada considérées comme non canadiennes	909	Sociétés constituées au Canada considérées comme non canadiennes
909	Autres services fédéraux		909	Activités indéterminées et imprécises	909	Activités indéterminées et imprécises

Division 9 — Finances, assurances et affaires immobilières — fin

Groupe 2 — Assureurs

721 Assureurs

Groupe 3 — Agences d'assurances et affaires immobilières

735 Agents d'assurances et agents immobiliers

737 Exploitants immobiliers

Division 10 — Services socio-culturels, commerciaux et personnels

Groupe 1 — Enseignement et services annexes

801 Jardins d'enfants et écoles maternelles

802 Écoles primaires et secondaires

803 Écoles des beaux-arts

804 Centres de formation professionnelle, écoles de métiers

805 Établissements d'enseignement postsecondaire non

806 Universités et collèges

807 Bibliothèques et musées

808 Enseignement et services annexes, n.c.a.

Groupe 2 — Services médicaux et sociaux

821 Hôpitaux

822 Établissements annexes de soins sanitaires

823 Cabinets de médecins et de chirurgiens

824 Cabinets de praticiens paramédicaux

825 Cabinets de dentistes

826 Services de diagnostic et de soins, n.c.a.

827 Services de santé divers

828 Organismes de bien-être

Groupe 3 — Organisations culturelles

831 Organisations culturelles

Groupe 4 — Divertissements et loisirs

841 Cinémas

842 Production et distribution de films

843 Salles de quilles et de billard

844 Clubs de golf et "country clubs"

845 Services de théâtre et autres spectacles

849 Services divers de divertissements et de loisirs

Groupe 5 — Services fournis aux entreprises

851 Bureaux de placement et services de location de per-

853 Services d'information

855 Services de sécurité et d'enquêtes

861 Bureaux de comptabilité

862 Services de publicité

863 Bureaux d'architecture

864 Bureaux d'études et services scientifiques

866 Études d'avocat et de notaire

867 Bureaux de conseil en gestion et en organisation

869 Services divers fournis aux entreprises

Groupe 6 — Services personnels

871 Cordonnettes

872 Salons de coiffure pour hommes et pour dames

873 Ménages

874 Blanchisseries et établissements de nettoyage à sec et

876 de presse (sauf en libre-service)

877 Libre-service de blanchissage et de nettoyage à sec

877 Pompes funèbres

879 Services personnels divers

Groupe 7 — Hébergement et restauration

881 Hôtels et motels

883 Pensions de famille et hôtels privés

Division 7 — Transports, communications et autres services publics — fin

Groupe 3 — Communications — fin

544 Réseaux de téléphonie

545 Réseaux de télégraphie et de câbles

548 Postes

Groupe 4 — Énergie électrique, gaz et eau

572 Énergie électrique

574 Distribution de gaz

576 Distribution d'eau

579 Autres services d'utilité publique

Division 8 — Commerce

Groupe 1 — Commerce de gros

602 Grossistes en produits agricoles

606 Grossistes en charbon et en coke

608 Grossistes en produits pétroliers

611 Grossistes en papier et articles en papier

612 Grossistes en marchandises diverses

614 Grossistes en alimentation

615 Grossistes en produits du tabac

616 Grossistes en médicaments et en produits de toilette

617 Grossistes en habillement et en mercerie

618 Grossistes en meubles de maison et accessoires d'a-

meublement

619 Grossistes en véhicules automobiles et accessoires

621 Grossistes en machines, matériel et fournitures élec-

triques

622 Grossistes en machines et matériel agricoles

623 Grossistes en machines et matériel, n.c.a.

624 Grossistes en quincaillerie, plomberie et matériel de

chauffage

625 Grossistes en métaux et produits métalliques, n.c.a.

626 Grossistes en sclages et matériaux de construction

627 Grossistes en déchets et matériaux de récupération

629 Grossistes, n.c.a.

Groupe 2 — Commerce de détail

631 Magasins d'alimentation

642 Magasins de marchandises diverses

652 Détaillants en pneus, accumulateurs et accessoires

654 Stations-service et postes d'essence

656 Détaillants en véhicules automobiles

658 Ateliers de réparation de véhicules automobiles

663 Magasins de chaussures

665 Magasins de vêtements pour hommes

667 Magasins de vêtements pour dames

669 Magasins de vêtements et de mercerie, n.c.a.

673 Quincailleries

676 Magasins de meubles et d'appareils ménagers

678 Ateliers de réparation de radios, de téléviseurs et d'ap-

pareils électriques

681 Pharmacies

691 Librairies et papeteries

692 Fleuristes

694 Bijouteries

695 Ateliers de réparation de montres et de bijoux

696 Magasins de vente de spiritueux, de vin et de bière

697 Débits de tabac

699 Détaillants, n.c.a.

Division 9 — Finances, assurances et affaires immobilières

Groupe 1 — Institutions financières

701 Banques et autres établissements de dépôts

703 Autres organismes de crédit

705 Agents de change et courtiers en valeurs mobilières

707 Sociétés d'investissement et sociétés à portefeuille

715 Bureaux canadiens de sociétés constituées au Canada

considérées comme non canadiennes

Division 3 — Industries manufacturières — fin		Groupe 10 — Industrie du papier et activités annexes		Division 5 — Industries manufacturières — suite	
				</	

CLASSIFICATION DES ACTIVITES ECONOMIQUES (1970)

N° de l'ind.	Division 1 - Agriculture	N° de l'ind.	Division 3 - Industries manufacturières
001	Groupes 1 - Fermes expérimentales et d'institution	101	Groupes 1 - Industries des aliments et boissons
003	Fermes expérimentales et universitaires	102	Industries de la viande et de la volaille
011	Groupes 2 - Fermes (sauf expérimentales et d'institution)	103	Industrie de la transformation du poisson
013	Fermes d'élevage spécialisées et mixtes	104	Industrie de la transformation de légumes
015	Fermes d'élevage spécialisées et mixtes	105	Industrie laitière
017	Fermes fruitières et maraîchères	106	Ménagerie et fabrication de céréales de table
019	Fermes fruitières et maraîchères	107	Fabrication d'aliments pour les animaux
021	Fermes de grandes cultures spécialisées et mixtes	108	Boulangerie et pâtisserie (fabrication)
031	Fermes d'élevage mixtes	109	Industries alimentaires diverses
033	Fermes de cultures diverses	110	Industrie des boissons
039	Groupes 3 - Services agricoles	111	Industrie du tabac
041	Services agricoles	112	Traitement du tabac en feuilles
043	Groupes 2 - Services forestiers	113	Fabrication de produits du tabac
045	Services forestiers	114	Industrie du caoutchouc et des produits en matière plastique
047	Groupes 1 - Pêche	115	Industrie des produits en caoutchouc
051	Pêche	116	Fabrication d'articles en matière plastique, n.c.a.
052	Groupes 2 - Services de pêche	117	Industrie du cuir
053	Services de pêche	118	Fabrication de produits en cuir
054	Groupes 3 - Chasse et piégeage	119	Fabrication de gants en cuir
055	Chasse et piégeage	120	Fabrication de chaussures
056	Groupes 4 - Mines (y compris broyage), carrières et puits de pétrole	121	Tanneries
057	Mines (y compris broyage), carrières et puits de pétrole	122	Fabrication de gants en cuir
058	Mines métalliques	123	Fabrication de valises, sacs à main et menus articles en cuir
059	Mines métalliques diverses	124	Industrie textile
061	Mines d'uranium	125	Filature et tissage de coton
062	Mines de quartz aurifère	126	Filature et tissage de la laine
063	Mines de fer	127	Fabrication de fibres, fils et tissus artificiels et synthétiques
064	Mines de gypse	128	Cordons et ficelles (fabrication)
065	Mines non métalliques diverses	129	Industrie du feutre et du traitement des fibres
066	Groupes 3 - Mines non métalliques (sauf mines de charbon)	130	Industrie des tapis, des carpes et de la moquette
067	Mines d'amiant	131	Industrie des articles en grosse toile et des sacs de coton et de jute
068	Tourbières	132	Industrie des accessoires en tissu pour l'automobile
069	Mines de gypse	133	Industries textiles diverses
070	Mines non métalliques diverses	134	Groupes 6 - Bonneterie
071	Groupes 4 - Carrières et sablières	135	Industrie des bas et chaussettes
072	Carrières	136	Bonneterie (sauf fabrication de bas et chaussettes)
073	Sablières et gravières	137	Industrie des vêtements pour hommes
074	Groupes 5 - Services miniers	138	Industrie des vêtements pour dames
075	Forage de puits de pétrole à forfait	139	Industrie des vêtements pour enfants
076	Autre forage à forfait	140	Industrie des articles en fourrure
077	Services miniers divers	141	Industrie des corsets et soutiens-gorge
078	Industrie des meubles de bureau	142	Industries diverses de l'habillement
079	Industrie des articles d'ameublement divers	143	Industrie du bois
080	Industrie des lampes électriques et des abat-jour	144	Industrie des plaques et de contre-plaques
081	Industrie des meubles de maison	145	Fabriques de portes, châssis et autres bois ouvrés
082	Industrie des articles d'ameublement divers	146	Fabriques de boîtes en bois
083	Industrie des articles d'ameublement divers	147	Industrie des cercueils
084	Industrie des articles d'ameublement divers	148	Industries diverses du bois
085	Industrie des articles d'ameublement divers	149	Industrie des articles d'ameublement divers
086	Industrie des articles d'ameublement divers	150	Industrie des articles d'ameublement divers
087	Industrie des articles d'ameublement divers	151	Industrie des articles d'ameublement divers
088	Industrie des articles d'ameublement divers	152	Industrie des articles d'ameublement divers
089	Industrie des articles d'ameublement divers	153	Industrie des articles d'ameublement divers
090	Industrie des articles d'ameublement divers	154	Industrie des articles d'ameublement divers
091	Industrie des articles d'ameublement divers	155	Industrie des articles d'ameublement divers
092	Industrie des articles d'ameublement divers	156	Industrie des articles d'ameublement divers
093	Industrie des articles d'ameublement divers	157	Industrie des articles d'ameublement divers
094	Industrie des articles d'ameublement divers	158	Industrie des articles d'ameublement divers
095	Industrie des articles d'ameublement divers	159	Industrie des articles d'ameublement divers
096	Industrie des articles d'ameublement divers	160	Industrie des articles d'ameublement divers
097	Industrie des articles d'ameublement divers	161	Industrie des articles d'ameublement divers
098	Industrie des articles d'ameublement divers	162	Industrie des articles d'ameublement divers
099	Industrie des articles d'ameublement divers	163	Industrie des articles d'ameublement divers
100	Industrie des articles d'ameublement divers	164	Industrie des articles d'ameublement divers

Ainsi, nous avons utilisé une technique originale mais assez rigoureuse pour obtenir un ensemble de regroupements industriels. Notre principal objectif était d'établir empiriquement une ventilation industrielle efficace et de décrire clairement les étapes nécessaires. Nous avons, dans une certaine mesure, atteint ici cet objectif. Force est maintenant de nous interroger à savoir si notre ventilation est raisonnable et si elle sera utile pour notre exercice d'estimation et nos services au secteur privé. La recherche d'une réponse à ces questions constitue la prochaine étape de l'analyse.

Enfin la, praticabilité d'une désagrégation industrielle, tant du point de vue de la possibilité d'obtenir des données que du point de vue des besoins des utilisateurs qui consultent les statistiques produites à partir desdites données, doit, comme toujours, être le principal facteur déterminant de la forme finale de la liste des catégories d'activité.

Conclusion

En somme, nous pouvons constater que l'analyse de grappes nous offre une ventilation des activités économiques empiriquement vérifiable, qui comprend quatre-vingt catégories formées à partir de 265 activités uniquement en fonction des différences entre les profils professionnels. En outre, l'analyse de grappes nous fournit un moyen de réduire la quantité d'informations d'une manière justifiable du point de vue méthodologique et, dans la plupart des cas, par simple raisonnement. Le fait qu'un grand nombre de grappes produites par l'analyse recoupent certains groupes de la C.A.E. ne pose pas forcément un problème parce que, dans la majorité des cas, les grappes constituent des ensembles logiques qui reflètent les conditions réelles du marché du travail et la façon dont la plupart des gens perçoivent le monde industriel. Pour cette raison et parce qu'il est possible de mesurer la proportion de la variation expliquée par les grappes et le degré de corrélation inter-grappes et intra-grappes, cette méthode d'analyse semble robuste.

Toutefois, tel que nous l'avons souligné plusieurs fois précédemment et bien que l'analyse de grappes permette d'introduire certaines catégories intéressantes qui autrement ne seraient qu'hypothétiques et sans appui empirique, cette technique constitue seulement un critère minimum pour une ventilation pratique des activités économiques. L'importance de chaque grappe telle que mesurée en fonction de la part de l'ensemble de la population active employée dans une branche d'activité, doit être prise en compte avant l'établissement d'une ventilation finale. Par ailleurs, la similarité des taux de croissance de l'emploi et le degré d'intégration industrielle implicite ont aussi un effet sur la divisibilité des activités.

travaux de maréchalerie et de soudure, et la grappe 71, qui réunit les cabinets de dentiste et les services de théâtre.

Administration publique

L'analyse de grappes a divisé ce secteur en trois parties. Les résultats s'écartent un peu de la C.A.E.

Puisqu'une grappe distincte a été formée pour la défense nationale (code 902), indépendamment de l'administration

fédérale. De plus, les administrations fédérale et provinciales sont regroupées -- ce qui semble assez logique -- dans la grappe 81. Enfin, le caractère unique de l'administration locale exige la création d'une grappe individuelle pour ce palier

d'administration. Une telle ventilation devrait mettre en évidence le fait que l'absence d'une division de ce secteur dans les soixante-neuf catégories initiales n'offrirait peut-être pas une représentation juste de la réalité.

activité économique et la plupart de ces résultats semblent intuitivement justes. Il serait en effet bien difficile de décider avec quelles autres activités il faut regrouper les cordonneries, les pompes funèbres, les services de sécurité, les services culturels, les salons de coiffure et les blanchisseries!

Quant aux grappes composées de plusieurs activités économiques, la plupart d'entre elles semblent "logiques".

Ainsi, le regroupement des services d'entretien de bâtiments comme la conciergerie et le nettoyage, avec les services d'hébergement des hôtels, des restaurants et des tavernes, peut être attribuable au fait qu'une grande part des établissements de ce genre possèdent leur propre personnel pour effectuer les travaux d'entretien. Le fait que deux activités nettement

divisée dans la C.A.E. -- la production de films cinématographiques (code 842) et la photographie et le développement de films photographiques (code 893) -- figurent ensemble dans la grappe 61 semble assez raisonnable, étant donné que ces deux activités se rapportent essentiellement au même produit, le film, sauf que dans un cas ce sont des images animées qui sont reproduites. En outre, l'analyse de grappes met les cabinets de médecin (codes 823 et 825) dans une grappe autre que celle des autres services médicaux (comme dans les soixante-neuf activités initiales), ce qui semble logique, et divise les services éducatifs en trois catégories, à savoir l'enseignement "non postsecondaire" (grappe 59), l'enseignement postsecondaire non universitaire ou collégial (grappe 62) et l'enseignement universitaire et collégial (grappe 73).

Enfin, on note deux groupements inhabituels: la grappe 63 qui comprend les écoles des beaux-arts (code 803) et les

L'habillage, la réparation de petits appareils électriques et, enfin, la réparation de montres et de bijoux. La ventilation du commerce de gros s'est faite en fonction de la distinction entre les industries primaires et non primaires. La grappe 52 comprend les entreprises qui vendent des produits issus du secteur primaire (comme, par exemple, les produits agricoles, le charbon, le coke et les sciages), tandis que la grappe 53 contient les entreprises qui vendent d'autres produits en gros.

Finances, assurances et affaires immobilières

L'analyse de grappes a produit une ventilation assez raisonnable de cette division industrielle. Comme on peut le constater au tableau des résultats, les services de nature financière tels que les banques et les courtiers en valeurs mobilières s'insèrent dans une même grappe (comme dans la C.A.E.), alors que les entreprises d'assurances et de biens immobiliers constituent une deuxième grappe. Il est intéressant de noter que l'analyse de grappes regroupe les assureurs (code 721) avec les agents d'assurances et les agents immobiliers (code 735) et les exploitants immobiliers (code 737). Bien que contraire aux indications de la C.A.E., ce résultat ne nous apparaît guère surprenant.

Services

Cette division pose un autre problème intéressant pour l'analyse de grappes puisqu'elle englobe un grand nombre d'activités hétérogènes dont beaucoup affichent des profils professionnels uniques. Une indication de cette complexité réside dans le fait que plusieurs grappes ne représentent qu'une seule activité économique. En effet, parmi les vingt-quatre grappes qui ont été formées, dix ne correspondent qu'à une seule

Commerce

Par contre, deux résultats imprévus figurent dans cet ensemble d'activités. Premièrement, les transports aériens et les services divers auxiliaires des transports sont regroupés; cela nous apparaît difficile à comprendre, à moins que les agences de voyages ne représentent la majorité des services divers auxiliaires. Enfin, il est surprenant de voir réunis les transports ferroviaires et les transports par eau.

postes; cette dernière activité est assez différente des secteurs de communications plus "traditionnels" comme le téléphone, la radio et la télévision. L'entreposage demeure dans une grappe à lui seul, comme dans la C.A.E.

Le commerce est une autre domaine dans lequel l'analyse de grappes a porté fruit. Six grappes ont été formées, alors qu'à l'origine il n'y en avait que deux parmi les soixante-neuf catégories initiales (travailleurs salariés et autres travailleurs) ou dans la C.A.E. (commerce de gros et de détail). Ceci semble être une ventilation plus raisonnable du commerce. La nécessité d'une meilleure ventilation du commerce a été admise dans la version de la classification des activités économiques publiée en 1980¹⁰, laquelle désagrége en seize catégories cette division industrielle. Cela dit, l'analyse de grappes préserve la distinction existante entre le commerce de détail et le commerce de gros puisque chaque grappe n'appartient qu'à une seule de ces deux grandes classes d'activités. Par exemple, l'analyse de grappes divise le commerce de détail en quatre grappes: le commerce de détail afférent aux véhicules automobiles (à l'exception des pneus), le commerce de détail non afférent aux véhicules automobiles (comme l'alimentation et

L'application de l'analyse de grappes à cette industrie a produit des résultats intéressants. Ainsi, on retrouve dans une seule grappe le déménagement et l'entreposage de biens usagés, la catégorie "autre camionnage", l'entretien de routes et de ponts et les autres services publics (qui comprennent l'enlèvement des ordures), ce qui apparaît raisonnable étant donné que toutes ces activités reposent sur l'utilisation du camion. Une autre aggrégation qui retient notre attention est celle des transports par autocar avec les "autres transports", qui incluent l'exploitation de cars d'excursion et le ramassage scolaire. Par ailleurs, il est encore plus intéressant de noter la fusion des transports par pipe-line et des services d'utilité publique (à l'exclusion des "autres services d'utilité publique", surtout parce que les transports par pipe-line concernent principalement la distribution physique de produits par pipe-line. On peut aussi signaler la formation de deux grappes distinctes pour l'exploitation de taxis et les

publics

Transports, entreposage, communications et autres services

classification précédente.

répartition en activités salariées et non salariées de la du bâtiment et des travaux publics a plus de sens que la assez logique. En effet, cette manière de diviser l'industrie canalisations de gaz et la construction de voies ferrées, semble englobe les travaux d'adduction d'eau, la construction de catégorie des autres travaux de construction (code 409), qui ponts et de voies publiques (code 406) soit regroupée avec la groupes susmentionnés. En outre, le fait que la construction de de bâtiments de tous genres", ce qui justifie la fusion des deux "font souvent sur place des travaux de réparation et d'entretien manuel de la C.A.E. (1970)⁹, les entrepreneurs spécialisés spécialisés (code 421) et le bâtiment (code 404). Selon le

L'industrie du bâtiment et des travaux publics pose elle aussi des problèmes, mais pour des raisons différentes. Comme il le sera démontré en détail dans le projet d'étude de l'industrie du bâtiment et des travaux publics, cette division comprend tout un éventail d'activités de construction qui exigent différentes combinaisons de compétences des travailleurs. Malheureusement, le recensement a divisé ce secteur en quatre sous-secteurs seulement, ce qui n'est sûrement pas assez précis pour bien décrire cette partie de l'économie. Quoi qu'il en soit, l'analyse de grappes faite à partir des distributions des professions a regroupé les entrepreneurs

Bâtiment et travaux publics

reflètent le mieux la réalité. d'activités de la C.A.E. ne sont pas nécessairement celles qui Toutefois, tel que précédemment mentionné, les catégories correspondait dans l'ensemble aux codes d'indice 2 de la C.A.E. différent de celui du système à soixante-neuf activités, qui présente ici pour les industries manufacturières est assez On note d'ailleurs que le schéma de groupement reflète la réalité. On constate aussi que l'analyse de grappes a mis la construction de navires et d'embarcations dans une catégorie à part, résultat dont les études sectorielles devront possiblement tenir compte. réfrigérateurs, d'appareils électriques, d'appareils d'éclairage et de fils électriques, et la grappe 23 qui groupe la fabrication de machines pour le bureau, de radioécepteurs, de téléviseurs, d'équipement de télécommunication et de matériel scientifique. Il est intéressant de noter que l'analyse de grappes a éliminé la fabrication des lampes électriques de la catégorie des meubles et des articles d'ameublement pour la regrouper avec la catégorie des appareils d'éclairage au sein de la grappe 15. On constate aussi que l'analyse de grappes a mis la construction de navires et d'embarcations dans une catégorie à part, résultat dont les études sectorielles devront possiblement tenir compte.

au niveau des professions. Toutefois, il est certainement normal, à première vue, de mettre en doute la logique d'une aggrégation des mines de charbon, des tourbières et des mines non métalliques diverses avec les mines métalliques d'or, d'uranium et de fer!

Industries manufacturières

Il s'est avéré que les résultats de l'analyse de grappes sont moins satisfaisants dans le cas des industries manufacturières et des industries des services que dans les autres domaines, en ce sens qu'il est difficile de produire des schémas logiques et faciles à expliquer. Ce problème tient sans doute au fait que les industries manufacturières englobent toute une gamme d'activités différentes pour lesquelles le profil des professions varie considérablement. Le fait qu'il existe plusieurs distributions uniques dans les industries

manufacturières complique le processus de répartition des activités dans une grappe appropriée et peut produire des groupes plutôt inattendus, comme on peut le voir au tableau 1. Prenons, par exemple, la grappe 16 qui regroupe les textiles et la bijouterie, la grappe 26 qui réunit les fabricants d'aéronefs, de fil métallique et d'outillage et les ateliers d'usinage, et la grappe 35 qui combine le secteur de la transformation du poisson avec les fabriques de chaussures.⁸

Malgré tout et bien qu'elles ne soient pas basées sur

un ordre de contiguïté similaire aux catégories de la C.A.E., plusieurs activités assument quand même un caractère logique lorsqu'on les examine d'un peu plus près. Mentionnons comme exemples la grappe 15 qui comprend les fabricants de lampes, de

8. Notons que les activités qui théoriquement ne semblent pas appartenir à la grappe dans laquelle elles ont été placées ont manifesté un faible degré de cohésion (corrélation) dans leur propre grappe ou une forte corrélation avec la grappe la plus proche, voire les deux. Ce genre de mesure n'est pas présentée ici, mais ces statistiques sont disponibles sur demande.

Ici encore l'analyse de grappes répartit une division industrielle, celle des forêts, en sous-divisions distinctes, à savoir l'exploitation forestière et les services forestiers, alors que le système précédent ne distinguait qu'un groupe d'activités salariées et un groupe d'activités non salariées. Il semble que le profil professionnel de l'exploitation forestière et celui des services forestiers, qui comprend les patrouilles, la lutte contre les incendies, les services de pépinières et le reboisement, soient très différents.

Chasse, pêche et piégeage

Compte tenu des différences entre les profils des professions dans cette branche, on obtient maintenant trois sous-divisions au lieu d'une seule. Il semblerait logique qu'il existe des différences entre les professions de l'industrie de la pêche, celles des services de pêche (qui englobent la pisciculture, les services d'inspection et de protection) et celles de la chasse et du piégeage. Toutefois, compte tenu de la forme sous laquelle les données sont recueillies et de la taille relative de ces trois activités, il est probable qu'on continue à produire des statistiques distinctes pour chacune de ces unités.

Mines

Dans l'industrie des mines, l'analyse de grappes à partir des profils des professions a eu pour effet de grouper des activités dont certaines semblent avoir peu de rapports entre elles. Quoi qu'il en soit, on a pu constater, en examinant la distribution des professions dans ces activités, que les industries regroupées affichent une structure semblable

grappes. Les chiffres soulignés qui figurent à la même ligne que le nom de la division (ou du secteur) montrent la proportion de la variation totale dans toutes les activités d'un secteur qui est expliquée par les grappes obtenues. Les chiffres en dessous des valeurs soulignées mesurent la proportion de la variation intra-grappes expliquée par les composantes calculées. On peut voir que la plupart de ces proportions sont supérieures à cinquante pour cent, ce qui signifie que la plus grande partie de la variation est expliquée par les grappes qui ont été formées. Les grappes où le pourcentage de variation expliquée est inférieur à cinquante se composent, comme nous le décrivons plus loin, d'activités qui semblent avoir peu de choses en commun mais qui sont tout de même regroupées. Ainsi, l'analyse de grappes offre non seulement des résultats qui sont compréhensibles intuitivement, mais aussi une indication quantitative du pouvoir explicatif des grappes obtenues.

Le tableau des résultats indique également les codes d'indice 3 de la C.A.E. pour les activités économiques regroupées dans chaque grappe. La définition de ces codes est donnée en annexe. Le tableau montre aussi la proportion de la population active incluse dans chaque grappe en 1971 et en 1981. Ces données ne sont pas examinées ici, mais sont plutôt présentées à titre de renseignement.

Agriculture

Une analyse de grappes appliquée aux profils des professions des deux industries agricoles -- les fermes et les services agricoles -- de la bande de recensement a révélé des différences prononcées entre elles. C'est là tout un changement par rapport à notre ancien système de 69 groupes qui divisait l'agriculture en une catégorie d'activités salariées et non salariées, respectivement.

62	Enseignement post-secondaire non universitaire	804, 805, 809	83.	0.43	0.54
63	Écoles des beaux-arts, maréchalerie et soudure	803, 896	52.	0.19	0.25
64	Bureaux d'architectes, bureaux d'études	863, 864	70.	0.71	0.99
65	Organismes de bien-être, ménages, services personnels divers	828, 873, 879	63.	1.71	1.89
66	Cordonneries	871	100.	0.04	0.02
67	Pompes funèbres	877	100.	0.09	0.07
68	Bureaux de comptabilité, bureaux de conseil en organisation des entreprises, services divers fournis aux entreprises	861, 867, 869	67.	0.69	0.54
69	Services de sécurité	855	100.	0.21	0.30
70	Services culturels	831	100.	0.59	0.51
71	Cabinets de dentiste, services de théâtre	825, 845	53.	0.30	0.45
72	Cinémas, salles de quilles, clubs de golf	841, 843, 844, 849	58.	0.78	0.91
73	Universités et collèges	806	100.	1.24	1.15
74	Services de publicité	862	100.	0.18	0.22
75	Salons de coiffure pour hommes et pour dames	872	100.	0.69	0.54
76	Services d'informatique	853	100.	0.06	0.26
77	Blanchisseries	874, 876	87.	0.52	0.34
78	Cabinets de médecine	823	100.	0.55	0.56
79	Bibliothèques et musées	807	100.	0.18	0.23

DIV. 11: ADMINISTRATION PUBLIQUE

80	Défense nationale	902	<u>88.</u>	100.	1.02
81	Administration fédérale et provinciale	909, 931, 991	80.	4.45	4.38
82	Administration locale	951	100.	1.85	1.98

Notes

1. Pour une explication des codes, voir l'annexe.
2. Les chiffres soulignés représentent la proportion de la variation totale des activités d'un secteur qui est expliquée par les diverses grappes; les chiffres présentés en-dessous mesurent la proportion de la variation intra-grappes qui est expliquée par chaque composante.
3. Ces chiffres indiquent la proportion de l'ensemble de la population active expérimentée à l'intérieur de chaque grappe selon les recensements de 1971 et de 1981, dans lesquels l'expérience était définie comme étant le fait d'avoir de l'expérience sur le marché du travail au cours des dix-huit mois qui ont précédé le mois de juin de l'année de recensement.

DIVISION 7: TRANSPORT, ENTREPOSAGE, COMMUNICATIONS ET AUTRES SERVICES PUBLICS									
39	Déménagement, camionnage, entretien de routes, autres services d'utilité publique	506, 507, 516, 579	70.	1.87	1.67				
40	Transports par autocar, autres transports	508, 509, 519	98.	0.49	0.59				
41	Transports par pipe-line, services d'utilité publique	515, 572-576	62.	1.08	1.03				
42	Entreposage	524, 527	68.	0.20	0.19				
43	Transports aériens, services divers auxiliaires des transports	501, 502, 517	61.	0.57	0.78				
44	Communications à l'exclusion des postes	543-545	63.	1.34	1.41				
45	Transports ferroviaires, transports par eau	503-505	49.	1.87	1.24				
46	Exploitation de taxis	512	100.	0.31	0.25				
47	Postes	548	100.	0.71	0.65				
DIVISION 8: COMMERCE									
48	Commerce de détail non afférent aux véhicules automobiles	631-652, 663-676, 681-694, 696-699	79.	8.92	8.69				
49	Commerce de détail afférent aux véhicules automobiles	654-658	63.	2.53	2.34				
50	Réparation de petits appareils électriques	678	100.	0.11	0.09				
51	Réparation de montres et de bijoux	695	100.	0.02	0.02				
52	Commerce de gros de produits issus du secteur primaire	602, 606, 608, 622, 626, 627	71.	1.25	1.21				
53	Commerce de gros de produits non issus du secteur primaire	611-621, 623-625, 629	60.	3.14	3.45				
DIV. 9: FINANCES, ASSURANCES ET AFFAIRES IMMOBILIERES									
54	Finances	701-707	65.	2.33	2.65				
55	Assurances, affaires immobilières	721-737	65.	2.15	2.52				
DIV. 10: SERVICES									
56	Bureau de placement, études d'avocat, syndicats ouvriers	851-866, 891, 899	71.	1.24	1.79				
57	Hébergement, entretien de bâtiments et d'habitations	881-886, 898	54.	4.58	6.05				
58	Services médicaux à l'exclusion des cabinets de médecin	821, 822, 824, 826, 827	60.	4.92	4.89				
59	Enseignement primaire et secondaire	801, 802	88.	5.21	4.38				
60	Location, réparations diverses	894, 895, 897	61.	0.26	0.34				
61	Développement de films : films cinématographiques, films photographiques	842, 893	63.	0.17	0.22				

16	Textiles, bijouterie	186-189, 392	57.	0.64	0.44
17	Cuir, produits en argile, abrasifs, chaux	172, 351, 357-359	48.	0.26	0.22
18	Première transformation des métaux	291-298	73.	1.44	1.14
19	Aliments à l'exclusion du poisson, de la boulangerie, des industries alimentaires diverses	101, 103-106, 109	46.	1.71	0.87
20	Portes, châssis, cerceuils, meubles	254, 258, 261-266	85.	0.74	0.84
21	Imprimerie, édition	286-289	65.	1.30	1.13
22	Articles de sport, enseignes, industries manufacturières diverses	393-399	73.	0.45	0.40
23	Machines pour le bureau, radiorécepteurs, téléviseurs, équipement de télécommunication, matériel scientifique	318, 334, 335, 391	59.	1.11	0.88
24	Ciment, béton, dérivés divers du pétrole	352, 354, 355, 369	67.	0.30	0.25
25	Industrie du papier et activités annexes à l'exclusion du papier de couverture asphalté	271, 273, 274	66.	1.55	1.21
26	Fil métallique, outillage, ateliers d'usinage, aéronefs	305, 306, 308, 321	54.	0.94	0.91
27	Construction de navires et d'embarcations	327, 328	89.	0.23	0.20
28	Boîte en bois, industries diverses du bois	256, 259	89.	0.12	0.13
29	Produits en caoutchouc, articles en matière plastique	162, 165	85.	0.57	0.57
30	Véhicules automobiles, carrosseries de camions	323, 324	88.	1.31	1.07
31	Asphalte, produits du pétrole, engrais, etc.	272, 365, 372, 373, 378	63.	0.62	0.58
32	Scleries, fabriques de placage	251, 252	77.	0.94	0.85
33	Tabac	151, 153	92.	0.11	0.07
34	Produits en pierre et articles de verre	353, 356	71.	0.20	0.15
35	Industrie de la transformation du poisson, fabriques de chaussures	102, 174	53.	0.61	0.68
36	Boulangerie, industries alimentaires diverses	107, 108	79.	0.95	0.75
<u>DIVISION 6: BATIMENT ET TRAVAUX PUBLICS</u>					
37	Bâtiment, entrepreneurs spécialisés	404, 421	83.	5.35	5.21
38	Construction de voies publiques, autres travaux de construction	406, 409	85.	1.43	1.06

(suite)

GROUPEMENTS DES ACTIVITÉS ÉCONOMIQUES PRODUITS PAR L'ANALYSE DE GRAPPES, POURCENTAGE DE LA VARIATION
EXPLIQUÉ ET PROPORTION DE LA POPULATION ACTIVE DANS CHAQUE ACTIVITÉ

Numéro de la grappe	Description de la grappe	Activités regroupées ¹ (codes de la CAF)	Variation expliquée ² (pourcentage)	Proportion de la population active ³ 1971 (pourcentage)
<u>DIVISION 1: AGRICULTURE</u>				
1	Fermes	001-019	<u>100.</u> 100.	5.95
2	Services agricoles	021	100.	0.12
<u>DIVISION 2: FORÊTS</u>				
3	Exploitation forestière	031	<u>100.</u> 100.	0.80
4	Services forestiers	039	100.	0.14
<u>DIVISION 3: CHASSE, PÊCHE ET PIÈGEAGE</u>				
5	Pêche	041	<u>100.</u> 100.	0.29
6	Services de pêche	045	100.	0.02
7	Chasse et piégeage	047	100.	0.01
<u>DIVISION 4: MINES</u>				
8	Mines métalliques, mines de charbon, tourbières, mines non métalliques diverses	051-061, 072, 079	78.	0.92
9	Pétrole, gaz naturel, forage	064, 096, 099	76.	0.45
10	Mines non métalliques	071, 073, 083, 087	76.	0.20
11	Autre forage à forfait	098	100.	0.04
<u>DIVISION 5: INDUSTRIES MANUFACTURIÈRES</u>				
12	Produits en métal, instruments aratoires, machines diverses et matériel ferroviaire	301-304, 309-315, 326, 329	67.	1.99
13	Gants, valises, textiles, habillement	175, 179, 187-249	80.	1.83
14	Industrie chimique à l'exclusion des fabricants de produits chimiques industriels	374-377, 379	78.	0.61
15	Lampes, équipement de réfrigération, appareils électriques, appareils d'éclairage, équipement électrique	268, 316, 331-333, 336-339	77.	0.86

(suite)

La colonne du tableau des résultats qui indique le pourcentage de la variation expliquée par les divers composantes permet de constater un des principaux avantages de l'analyse de

Cette section présente les résultats de l'analyse de grappes faite à partir des profils des professions des 265 activités économiques du fichier de recensement, selon des catégories générales d'activités. Les groupements ainsi produits figurent au tableau ci-inclus. Avant de faire quelques brèves observations sur chaque division, rappelons encore une fois que ces résultats ne constituent aucunement un schéma d'agrégation définitif. Des facteurs tels que la taille relative des grappes, les restrictions imposées par les données disponibles, les rapports étroits qui existent entre certaines activités ainsi que les besoins des utilisateurs en matière de recueils de statistiques, exigent certes des modifications de la structure industrielle établie empiriquement. Bien que notre intérêt porte essentiellement sur les professions, ce n'est certainement pas le seul critère qui a motivé notre décision de considérer des activités économiques en tant qu'unités distinctes ou composantes d'un groupe.

b) Résultats finals de l'analyse de grappes

analyse exigerait énormément d'espace mémoire pour le traitement par S.A.S. Par ailleurs, certaines grappes, tel qu'il a été mentionné plus haut, pourraient être dépourvues d'une base solide dans le monde réel et s'avérer difficiles à défendre rationnellement. Bref, bien que les grappes soient justifiables statistiquement, il peut être difficile de trouver une explication logique de leur existence. Pour ces raisons, l'analyse de grappes doit se faire à partir de sous-ensembles de données déterminées par la forme des renseignements existants, c'est-à-dire selon les divisions de la C.A.E.

IV. Résultats

a) Résultats préliminaires

Abstraction faite des réserves mentionnées plus haut, une analyse préliminaire de l'ensemble des coefficients de corrélation traités au préalable par A.P.L. a révélé des résultats intéressants. Ainsi, l'application d'une analyse de grappes à vingt-neuf secteurs dissemblables a produit des grappes qui confirment bon nombre de nos intuitions relatives aux groupes avec lesquels ces secteurs seraient associés. Les grappes formées étaient en réalité une approximation des catégories de la C.A.E. L'intérêt de ce résultat tient au fait qu'une analyse effectuée seulement en fonction des (différences entre les) profils des professions semble ranger les activités économiques dans un schéma de classification qui reproduit un système (la C.A.E.) basé sur le produit, les intrants et le processus de production.⁷ Il serait en effet intéressant d'appliquer une analyse de grappes à toutes les 265 activités, une à la fois, pour observer à quel point les résultats correspondent à la C.A.E. Ce serait là une manière de faire table rase, pour ainsi dire; l'analyse se ferait alors indépendamment des contraintes imposées par certaines divisions logiques et sans biais à priori.

Même si une analyse de grappes des 265 activités économiques serait gratifiante sur le plan intellectuel, le projet ne serait peut-être pas pratique. D'abord, une telle

7. Pour une description détaillée des critères de division des activités économiques dans la C.A.E., voir Bureau fédéral de la statistique, Classification des activités économiques (no 12-501 au catalogue, Information Canada, Ottawa, décembre 1970, p. 12, 13).

Ainsi, nous devons également examiner la taille relative des grappes produites par l'analyse.

Tous ces aspects sont abordés tour à tour après la description des résultats des analyses de grappes. Ils ne sont ici mentionnés que dans le cadre d'un aperçu et pour souligner le fait que la similarité des structures professionnelles n'est pas le seul critère de groupement des activités économiques.

dans une autre (et vice versa).⁶ Une simple application de l'analyse de grappes ne permettrait pas, bien entendu, de déceler ces effets et c'est là un facteur qu'il faut prendre en compte dans l'examen des premiers résultats.

En outre, il est probable qu'une agrégation industrielle qui convienne à un moment particulier ne corresponde pas à la réalité d'une autre période à cause de l'évolution des conditions économiques et technologiques. Il serait peut-être intéressant de faire une analyse de grappes à partir des coefficients de 1971 codés de manière semblable à ceux de 1981. Une comparaison des grappes d'activités formées pour 1971 et 1981 ferait ressortir les changements des structures industrielles, lesquels reflèteraient les différences entre ces deux périodes au chapitre de l'économie et de la technologie.

Un dernier point qu'il importe de soulever ici concerne la taille des activités qui composent une grappe. Quels que soient les groupements produits par l'analyse, on doit tenir compte de l'importance relative de chaque catégorie d'activité. Autrement dit, nous devons nous demander si toutes les grappes sont suffisamment grandes pour justifier la création de groupes distincts. Il ne vaut pas la peine de calculer des prévisions de l'emploi pour une activité économique relativement peu importante quant à la part de l'ensemble de la population active qu'elle occupe, et ce même si sa structure professionnelle est unique. On inclurait plutôt cette activité économique dans la grappe avec laquelle elle a le plus de similarités.

6. Nous pensons plus particulièrement aux activités économiques implicitement intégrées comme, par exemple, l'industrie des forêts, dont une grande partie des ventes dépend des secteurs des produits forestiers et des pâtes et papiers. Dans les cas de ce genre, l'évolution de la production et, par conséquent, de l'emploi dans une activité économique devrait être synchronisée avec l'évolution de la production dans une autre.

Un autre point qu'il faut faire valoir se rapporte au fait que l'analyse de grappes soit non-paramétrique. D'abord et avant tout, ceci rend difficile l'évaluation rigoureuse de la signification statistique de cette méthode. Cependant, l'analyse de grappes peut quand même offrir une indication quantitative assez robuste des similarités pouvant exister entre les activités économiques.

En plus des limitations statistiques, il faut aussi

signaler une difficulté sur le plan théorique. Il est possible que l'analyse théorique conduise à une agrégation des activités économiques qui ne cadre pas avec la conception que les employeurs et les syndicats ont du marché du travail. Si notre schéma justifié empiriquement ne concorde pas avec l'organisation réelle du marché et des transactions entre employeurs et travailleurs, nous n'aurons vraiment pas gagné grand-chose. Voilà pourquoi nous mènerons notre analyse tel

qu'initialement prévu, en limitant les analyses de grappes successives à l'intérieur de certains groupes "naturels" d'activités économiques.

Un autre élément à souligner tient au fait que les activités économiques peuvent être semblables non seulement en raison du profil des professions, mais aussi à cause du taux de croissance de l'emploi (même si la base démographique n'est pas la même). Il n'est pas nécessaire d'effectuer des calculs distincts pour deux activités économiques qui croissent ou décroissent en même temps, parce qu'une moyenne pondérée des deux profils professionnels nous fournirait des renseignements équivalents à un coût moins élevé. On trouve un exemple précis dans le cas des activités économiques intégrées verticalement, où l'essor d'une activité dépend étroitement de la situation

L'application de l'analyse de grappes à la matrice de transformation industrie/profession nous conduira, au pire, à une justification empirique du groupement actuel des activités économiques. Jusqu'à présent, nous avons utilisé les 69 groupes d'activités économiques du PPC, quoiqu'il n'y ait jamais eu une explication claire justifiant l'utilisation de cette structure de préférence à une autre. Le degré d'agrégation des activités est presque déterminé à l'avance par certaines caractéristiques réelles des données dans certains secteurs, mais cela n'explique pas l'origine des catégories d'activité dans tous les secteurs. D'ores et déjà, nous aurons au moins une confirmation empirique de ce système.

Il convient d'émettre certaines réserves relatives à la théorie et la méthode que nous avons appliquées. D'abord, quel que soit le niveau de désagrégation industrielle prescrit par nos critères théoriques, il faut se rappeler qu'il existe des limites à la finesse des détails qu'on peut extraire d'une source de données (sans parler des modèles de prévisions) en ce qui a fait à la production, l'emploi et d'autres variables. Même si nous pouvons établir des relations fonctionnelles qui expliquent l'évolution temporelle de l'emploi dans diverses activités économiques, il est peu probable que nous puissions obtenir des prévisions relatives à plusieurs des variables indépendantes. Autrement dit, un modèle macro-économique comme TIM peut seulement produire des prévisions à un niveau assez global, ce qui rend une analyse détaillée impossible dans bien des cas. De toute manière, il peut être utile et intéressant en soi de découvrir empiriquement quelles activités économiques diffèrent en fonction du profil des professions.

Au début, VARCLUS place toutes les variables à l'intérieur d'une seule composante; les étapes suivantes sont par la suite exécutées itérativement:

(i) La grappe dont la proportion de la variation expliquée par sa composante est la plus faible est sélectionnée pour être divisée.

(ii) La grappe choisie est divisée suite au calcul et à la rotation des deux premières composantes principales, après quoi les variables sont insérées dans la grappe avec laquelle elles ont la plus grande corrélation.

(iii) Les variables sont ensuite réparties de nouveau dans des grappes pour maximiser la variation expliquée par les composantes des grappes.

Le traitement se termine dès que le critère qui spécifie que chaque grappe doit avoir une racine caractéristique supérieure à un, est satisfait. Ceci signifie qu'une seule composante d'une grappe explique une part plus grande de la variation que la variation normalisée "habituelle". Autrement dit, c'est à ce point qu'est réalisée la structure optimale empirique des grappes.

L'un des avantages de l'analyse de grappes est que l'utilisateur ne se heurte pas aux complexités de l'appareil théorique, contrairement à l'analyse de régression qui comporte des tests et des techniques qui s'avèrent souvent trop compliquées pour les non-spécialistes. Parce qu'elle est non-paramétrique, l'analyse de grappes s'applique aisément et produit des résultats qui sont faciles à interpréter intuitivement et à comprendre.

Pour nos besoins, nous avons employé le logiciel S.A.S. (Statistical Analysis System), lequel est muni d'un ensemble complet de procédés pour l'analyse de grappes. Nous avons choisi le procédé VARCLUS, qui exécute une analyse de variables (les activités économiques) et crée des grappes disjointes, ce qui correspond exactement au résultat recherché.

VARCLUS, comme la plupart des techniques d'analyse de grappes, repose essentiellement sur un critère de formation de grappes qui est en fait la formule de la dispersion (variation) totale de l'ensemble de données. La variation totale se compose de deux éléments, la variation intra-grappes et la variation inter-grappes. En général, l'analyse de grappes essaye de réduire au minimum la variation intra-grappes (ce qui revient à maximiser la variation inter-grappes) et répartit les variables dans des grappes en fonction de ce critère. Quand les données d'entrée sont des coefficients de corrélation, comme dans le cas présent, on peut simplement considérer ces derniers comme des mesures normalisées de la variation. Pour réduire le temps et le travail nécessaire, VARCLUS calcule et utilise des composantes qui sont des combinaisons pondérées indépendantes des variables qu'elles représentent, pour ensuite "maximiser, pour l'ensemble des grappes, la somme de la variation expliquée par les composantes des grappes".⁵

5. Pour plus de détails, voir le manuel S.A.S. User's Guide: Statistics, Op. Cit., p. 462-473 ou B. Everitt, Op. Cit.

variables à une dimension plus raisonnable sans perte importante de renseignements. Dans le cas que nous étudions ici, la ventilation la plus détaillée qui soit possible -- composée de 265 activités économiques distinctes -- serait trop désagrégée pour être pratique aux utilisateurs de tels renseignements. En outre, il serait tout simplement illogique d'aller dans le monde réel muni d'informations indiquant, par exemple, que les fabricants de savon devraient être regroupés avec les fabricants de ciment. Il faut donc user d'un minimum de jugement qualitatif et rationnel dans l'évaluation des conclusions de notre analyse.

Au cours des dernières années, l'utilisation de l'analyse de grappes s'est largement répandue dans divers domaines. Cette expansion multidisciplinaire a entraîné un foisonnement des méthodes d'analyse fondées sur différentes définitions de la notion de grappe et de similarité entre objets. Par exemple, il peut y avoir des grappes complètement disjointes formées d'unités qui entrent dans une grappe particulière et non ailleurs, ou des grappes qui se recoupent parce que les unités peuvent appartenir à plus d'une grappe. L'unité qui est incluse dans une grappe peut également être soit une variable (dans ce cas-ci, l'industrie), soit une observation (la profession). Il est enfin possible d'utiliser différentes mesures de la distance ou de la similarité entre les unités comparées. Un exemple applicable au premier cas serait celui de la matrice euclidienne, et un applicable au second le coefficient de corrélation.⁴

4. Pour davantage d'explication relatives à ces mesures et à d'autres, voir Everitt, Op. cit., pp. 49-59. Le coefficient de corrélation a été critiqué à titre de mesure de la similarité entre la forme des profils qu'aux valeurs elles-mêmes, "du fait que ce coefficient soit égal à un chaque fois que deux profils sont parallèles, peu importe la distance entre deux points" ((traduction libre) Everitt, Op. cit., p. 53). Ce coefficient est donc convenable ici puisque, comme on l'a souligné plus haut, c'est justement la distribution des professions dans les activités économiques qui nous intéresse, (d'autant plus que les données ont toutes été standardisées à l'échelle industrielle).

II. Analyse de grappes: une brève description

Avant d'aller plus loin, il semble opportun de résumer exactement ce qu'est l'analyse de grappes et de décrire les

avantages qu'elle présente. En un mot, l'analyse de grappes est un terme qui s'applique à un éventail de techniques qui ont pour objet de grouper les variables ou les observations d'un ensemble de données en un nombre plus restreint de grappes individuelles. Les composantes d'une grappe doivent, selon une certaine

définition, être semblables, tandis que chaque grappe doit se distinguer de toutes les autres. L'analyse de grappes vise donc à "placer des unités dans des ensembles ou des grappes qui naissent des données elles-mêmes et non d'une définition a priori, de sorte que les objets d'une grappe particulière

attendent une ressemblance générale dans un certain sens, alors que les unités de grappes différentes soient dissemblables"². En d'autres termes, l'objectif est de maximiser la similarité à l'intérieur de chaque grappe tout en la minimisant d'une grappe à l'autre.

Mis à part les cas où il s'agit simplement de résumer les données, l'analyse de grappes devrait produire un ensemble de grappes constituées de manière plus ou moins logique. La structure des grappes doit être "si évidente qu'il aurait été possible de la reconnaître, du moins dans les analyses faites sur une petite échelle, sans recours à des techniques mathématiques compliquées et sans définition précise du terme grappe"³. En fait, le but de la méthode est de trouver des grappes "naturelles" et de réduire en même temps le nombre de

2. S.A.S. User's Guide: Statistics, édition 1982, S.A.S. Institute Inc., Cary, N.C., Etats-Unis, page 417.
3. Everitt, Brian, Cluster Analysis, 2^e édition, Heinemann Educational Books, Londres, Royaume-Uni, 1980, page 7.

complication; il a donc fallu envoyer la forme complète de la matrice des 265 x 511 coefficients à I.P. Sharp pour le traitement par A.P.L. Le grand avantage d'A.P.L. est sa souplesse pour la manipulation des matrices.

Le traitement ci-dessus des enregistrements nuls au chapitre des professions qui n'entrent pas dans certaines activités économiques peut se justifier de la façon suivante. Nous sommes d'avis que la classification type des professions a été conçue afin d'englober toute la gamme des métiers qui existent dans l'économie canadienne, mais non pour s'appliquer parfaitement à toutes les activités économiques prises individuellement. Beaucoup de catégories de professions ne cadrent pas du tout avec la structure de l'emploi dans de nombreuses activités économiques. Par conséquent, c'est avec une certaine assurance que nous avons supprimé les comparaisons portant sur deux activités économiques pour lesquelles ne figurait ni dans l'une ni dans l'autre, une profession donnée. En revanche, dans les cas où une profession existait dans une activité économique mais non dans la deuxième, il a fallu élargir la liste des professions de cette dernière afin que le nombre de professions soit uniforme partout. Cela signifie que, pour permettre une comparaison dans tous les cas de cette nature, nous avons inséré la valeur zéro (indiquant une présence négligable) pour les activités économiques où une profession n'existait pas, sans modifier les valeurs non nulles qui existaient déjà.

économiques, nous n'avons obtenu qu'une seule grappe. Lorsque nous avons choisi de produire un minimum de cinq grappes, nous n'en avons obtenu que cinq; de même, quand nous avons spécifié un minimum de dix grappes, l'analyse a produit dix grappes, pas une de plus. De tels résultats semblent donc mettre en doute la valeur de notre analyse puis que nous faisons face exactement au problème qu'on voulait éviter: le choix arbitraire à priori d'un nombre fixe d'activités économiques.

Malgré tout, les résultats des analyses décrites plus haut ont semblé "logiques" en ce sens qu'ils reproduisaient assez fidèlement les groupements de la C.A.E.. Cela tendrait à prouver que l'analyse de grappes est une méthode assez robuste. Nous avons donc conservé l'espoir que cette méthode puisse conduire à une ventilation industrielle qui soit empiriquement justifiable, une fois résolu le problème des coefficients nuls.

La difficulté qui se pose est que le seul élément de S.A.S. qui ressemble le plus à ce dont nous avons besoin est une option de "suppression par paires" pour les valeurs manquantes (les valeurs nulles sont considérées comme manquantes). Quand on compare deux variables, cette option élimine une observation de l'analyse lorsque la valeur de l'une ou encore des deux variables est manquante. Pour nos besoins, toutefois, ce n'est que lorsque les deux variables font défaut que nous souhaiterions supprimer une observation. Autrement dit, nous voulons éliminer les cas (qui sont nombreux dans notre fichier) où deux activités économiques ne contiennent pas une même profession, sans rejeter les comparaisons portant sur les professions qui figurent dans l'une et/ou l'autre des deux activités économiques comparées. Malheureusement, tel que mentionné plus haut, S.A.S. ne peut venir à bout de cette

dans une activité économique. Enfin, on a "transposé" ce fichier de 265 x 511 enregistrements sur un troisième fichier ne contenant que 265 enregistrements, chacun de dimension 511 (professions).

La mesure acceptée de la proximité des unités dans une analyse de grappes est le coefficient de corrélation. On a donc calculé le degré de corrélation d'une activité économique à une autre en fonction de l'unité d'observation "profession". La matrice des corrélations ainsi produite a été jugée nécessaire afin de permettre des économies d'espace et de temps et d'accroître la flexibilité du traitement. Premièrement, la matrice que nous avons mise sur fichier compte 265 x 265 cases au lieu de 265 x 511 cases, et deuxièmement, la matrice de similarités (corrélations) utilisée dans l'analyse de grappes n'a pas besoin d'être recalculée à chaque fois qu'une analyse de grappes est effectuée pour différentes combinaisons d'activités économiques.

Cependant, une fois les coefficients de corrélation calculés à partir de cet ensemble complet de données, nous avons découvert que la plupart des coefficients avaient des valeurs élevées et proches l'une de l'autre. Ce résultat est sans doute attribuable au grand nombre de correspondances nulles qui existent entre les activités économiques au chapitre des professions: celles-ci prédominent à ce point dans les calculs qu'elles submergent les comparaisons non nulles et introduisent un biais dans les résultats. Pour cette raison, nous doutons que des techniques factorielles comme l'analyse de grappes puissent distinguer sans difficulté les différentes activités économiques. En effet, quand nous avons exécuté l'analyse de grappes de S.A.S. avec un groupe d'environ cinquante activités

Le fichier de recensement combiné d'origine ne comprend des enregistrements que pour chaque profession qui existait en 1981 à l'intérieur des classes individuelles d'activité économique d'indice 3 (il n'y a pas d'enregistrement pour les professions non requises dans une activité particulière). Ces enregistrements contiennent des chiffres sur la population active occupée et expérimentée, qui sont exprimés en valeur absolue et en pourcentage de l'emploi dans chaque activité économique pour 1971 et 1981. Même si la suppression des enregistrements de valeur nulle pour une profession donnée permet indéniablement de réduire les coûts de stockage des données, elle pose des problèmes pour la plupart des analyses statistiques. En effet, comme chaque activité économique exige différentes combinaisons de professions, il n'y aura pas une correspondance parfaite entre les professions des diverses activités, ce qui est essentiel pour des techniques de comparaison telles que l'analyse de grappes.

Il a donc fallu faire procéder à différentes manipulations afin de remanier ce fichier, dont la structure était conçue en fonction des économies d'espace. On a d'abord créé un fichier composé d'environ 136 000 enregistrements en blanc (265 activités économiques multipliées par 511 professions), ce qui représente le nombre maximum d'enregistrements possibles (compte tenu que chaque enregistrement porte un code d'activité économique et un code de professions). Ensuite, on a comparé le fichier de recensement combiné (décrit plus haut) avec ce nouveau fichier. A chaque fois qu'il y avait correspondance des codes, la part de la population active expérimentée en 1981 (la seule donnée dont nous avions besoin) a été copiée dans la zone correspondante du nouveau fichier, ce qui a eu pour effet de laisser des blancs là où une profession donnée n'existait pas

Utilisation de l'analyse de grappes pour déterminer le niveau
optimal de désagrégation industrielle:
Rapport d'activité

Introduction

Nous présentons ici les résultats d'un projet visant à établir une ventilation pratique des activités économiques sur une bonne base empirique.¹ Depuis sa création, le S.P.P.C. repose sur les soixante-neuf activités économiques du système qu'il a remplacé, le PPC. Cette répartition a sans doute été choisie pour des raisons logiques, mais ces arguments n'ont jamais été consignés dans un document. Voilà justement ce que propose de faire la présente étude. Comme ce sont avant tout les professions qui nous intéressent dans le S.P.P.C., notre objet est d'élaborer une désagrégation optimale des activités économiques à partir d'une comparaison empirique de la distribution des professions dans chaque activité.

La première section résume les étapes de préparation du fichier de recensement initial sous une forme convenant à l'analyse empirique. Suit un court exposé de la technique utilisée plus loin -- l'analyse de grappes -- et de ses avantages, tandis que la section III énumère certaines réserves sur le plan méthodologique et statistique. La section suivante présente et évalue les résultats d'une analyse de grappes faite à partir des coefficients des professions dans les industries classées dans le fichier de recensement. Une brève conclusion est donnée à la fin du rapport.

1. Ce document est essentiellement la suite d'une première description de ce projet, Le niveau de désagrégation industrielle dans le cadre du S.P.P.C. -- un document de travail, mars 1963, document présenté à la conférence technique d'avril 1983.

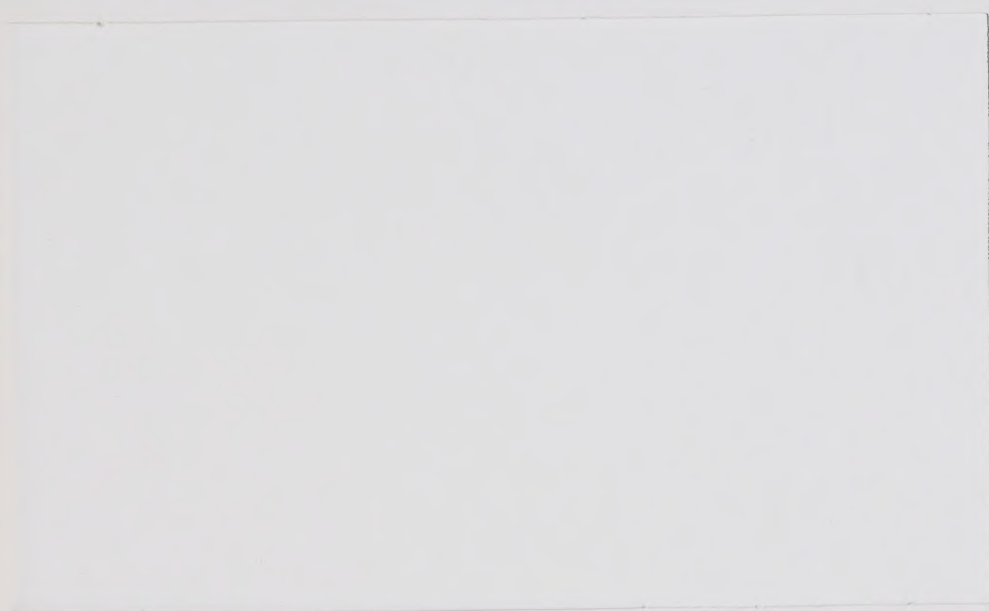
SYSTEME DE PROJECTION
DES PROFESSIONS AU CANADA
(SPPC)

LA DETERMINATION DU
NIVEAU OPTIMAL DE
DESAGREGATION INDUSTRIELLE -
RAPPORT D'ETAPE

PAR: NORM LECKIE
POLITIQUE STRATEGIQUE
ET PLANIFICATION
DIRECTORAT DES PERSPECTIVES
DU MARCHE DU TRAVAIL ET DE
L'ANALYSE STRUCTURELLE

NOVEMBRE 1983

WH-3-408B



LA DETERMINATION DU
NIVEAU OPTIMAL DE
DESAGREGATION INDUSTRIELLE -
RAPPORT D'ETAPE
PAR: NORM LECKIE
POLITIQUE STRATEGIQUE
ET PLANIFICATION
DIRECTORAT DES PERSPECTIVES
DU MARCHE DU TRAVAIL ET DE
L'ANALYSE STRUCTURELLE

Système de projections des professions au Canada



